The Bell System Technical Journal

October, 1934

An Extension of the Theory of Three-Electrode Vacuum Tube Circuits

By S. A. LEVIN and LISS C. PETERSON

The relations between input voltage and output current of the three-electrode vaccum tube are discussed when arbitrary feedback is present between grid and plate circuits. Fundamental assumptions are that the amplification factor is constant and conductive grid current absent. The relations developed in the present paper are generalizations of those given by J. R. Carson in I. R. E. Proc. of 1919, page 187. The use of the theory is illustrated by application to a simple modulator circuit. The numerical calculations in this case indicate that neglecting the effects of interelectrode tube capacitances may introduce serious errors.

Introduction

THE relations between input voltage and output current of the three-electrode vacuum tube when connected to impedances in both input and output circuits have been the subject of several papers. One of the first more extensive treatments of this problem was given by J. R. Carson,[1] using a method of successive approximations. The theory was further extended by F. B. Llewellyn,[2] E. Peterson and H. P. Evans,[3] and J. G. Brainerd.[4] The theories given by these authors did not take into account any feedback between input and output circuit except in the first approximation.

The aim of the present paper is to extend the theory of the three-electrode vacuum tube to include the effects of feedback between input and output circuits not only in the first but also in the second and higher approximations. The assumptions underlying Carson's treatment, constancy of the amplification factor and absence of conductive grid current, will be maintained. The extension of the present theory to such cases as treated by Llewellyn, Peterson-Evans and Brainerd still remains to be done.

[1] J. R. Carson: *I. R. E. Proc.*, April, 1919, page 187.
[2] F. B. Llewellyn: *B. S. T. J.*, July, 1926, page 433.
[3] E. Peterson and H. Evans: *B. S. T. J.*, July, 1927, page 442.
[4] J. G. Brainerd: *I. R. E. Proc.*, June, 1929, page 1006.

Theory

Let us consider the circuit arrangement shown in Fig. 1, where Z_1, Z_2 and Z_3 are linear impedances which may include interelectrode admittances. The impressed variable electromotive forces whose instantaneous values are denoted by E_g and E_p are in series with the impedances Z_g and Z_p, respectively. In the absence of these electromotive forces direct currents and voltages are established in the circuit due to constant grid and plate electromotive forces. With the variable electromotive forces impressed incremental currents and voltages are produced. The instantaneous values of these incremental voltages are indicated on Fig. 1 by g, e, v and p. The incremental plate current is J. The positive directions of these quantities are given by the directions of the arrows.

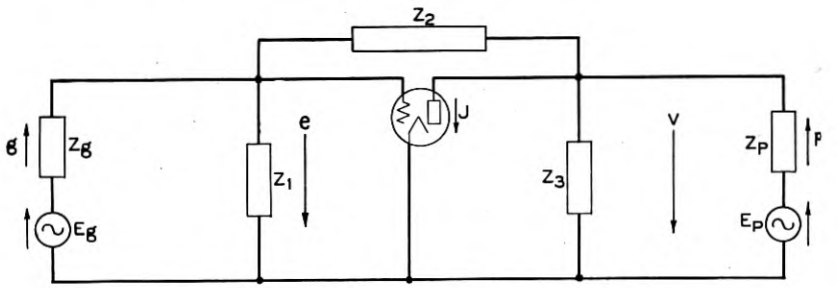


Fig. 1—Three-electrode vacuum tube and circuit.

We will now make two restrictive assumptions: first that the grid is never positive so that conductive grid current is absent, and second that the amplification factor μ is constant.

The basis for the analysis is given by the characteristic tube equation:

$$I = f\left(E_c + \frac{E_b}{\mu}\right), \tag{1}$$

where I is the total instantaneous current flowing from plate to filament; E_c is the total instantaneous potential difference between grid and filament and E_b the total instantaneous potential difference between plate and filament. μ is the amplification factor. The relation between the increments e, v and J is given by the following equation:

$$J = P_1(\mu e + v) + P_2(\mu e + v)^2 + \cdots + P_n(\mu e + v)^n + \cdots, \tag{2}$$

where

$$P_m = \frac{1}{m!}\frac{\partial^m I}{\partial E_b^m}$$

and has to be evaluated at the operating point.[1]

We have further:

$$E_g = g + e, \qquad E_p = p + v. \tag{3}$$

The equations (3) are obtained by applying the circuital laws to the network external to the tube.

We now proceed to a solution of equations (2) and (3) by means of a method of successive approximations. Let

$$J = \sum_1^\infty J_i, \qquad g = \sum_1^\infty g_i, \qquad e = \sum_1^\infty e_i, \tag{4}$$
$$p = \sum_1^\infty p_i, \qquad v = \sum_1^\infty v_i$$

and let us define the relations between the terms in the series (4) as follows:

$$J_1 = P_1(\mu e_1 + v_1), \qquad Eg = g_1 + e_1, \qquad E_p = p_1 + v_1, \tag{5}$$

$$J_2 = P_1(\mu e_2 + v_2) + P_2(\mu e_1 + v_1)^2, \tag{6}$$
$$0 = g_2 + e_2, \qquad 0 = p_2 + v_2,$$

$$J_3 = P_1(\mu e_3 + v_3) + 2P_2(\mu e_1 + v_1)(\mu e_2 + v_2) + P_3(\mu e_1 + v_1)^3, \tag{7}$$
$$0 = g_3 + e_3, \qquad 0 = p_3 + v_3,$$

$$J_4 = P_1(\mu e_4 + v_4) + P_2(\mu e_2 + v_2)^2 + 2P_2(\mu e_1 + v_1)(\mu e_3 + v_3) + 3P_3(\mu e_2 + v_2)(\mu e_1 + v_1)^2 + P_4(\mu e_1 + v_1)^4, \tag{8}$$
$$0 = g_4 + e_4, \qquad 0 = p_4 + v_4,$$

and so forth for subsequent terms.[5]

If we now let

$$R_0 = \frac{1}{P_1}, \tag{9}$$

[1] Loc. cit.

[5] The procedure of finding these equations is as follows: By substituting the first term in each of the series (4) into (2) and (3) and neglecting all terms higher than the first order equations (5) are obtained. By substituting the first two terms in each of the series (4) into (2) and (3), and neglecting terms of higher order than the second and by noting (5) equations (6) are found and so on for the remaining equations.

where R_0 is the internal resistance of the tube, equations (5), (6), (7) and (8) may be rewritten as:

$$R_0J_1 - v_1 = \mu e_1, \qquad E_g = g_1 + e_1, \qquad E_p = p_1 + v_1, \tag{10}$$

$$\begin{aligned} R_0J_2 - v_2 &= \mu e_2 + R_0P_2(\mu e_1 + v_1)^2, \\ 0 &= g_2 + e_2, \qquad 0 = p_2 + v_2, \end{aligned} \tag{11}$$

$$\begin{aligned} R_0J_3 - v_3 &= \mu e_3 + 2R_0P_2(\mu e_1 + v_1)(\mu e_2 + v_2) + R_0P_3(\mu e_1 + v_1)^3 \\ 0 &= g_3 + e_3, \qquad 0 = p_3 + v_3, \end{aligned} \tag{12}$$

$$\begin{aligned} R_0J_4 - v_4 &= \mu e_4 + R_0P_2(\mu e_2 + v_2)^2 + 2R_0P_2(\mu e_1 + v_1)(\mu e_3 + v_3) \\ &\quad + 3R_0P_3(\mu e_2 + v_2)(\mu e_1 + v_1)^2 + R_0P_4(\mu e_1 + v_1)^4, \\ 0 &= g_4 + e_4, \qquad 0 = p_4 + v_4, \end{aligned} \tag{13}$$

and so forth.

Equations (10) to (13) admit of simple physical interpretations. Referring first to equations (10) it is clear that the equivalent circuit corresponding to Fig. 1 for first order quantities is given by Fig. 2. Similarly Fig. 3 is the equivalent circuit of Fig. 1 for second order effects and Fig. 4 for third order effects. Higher order effects correspond to similar circuits.

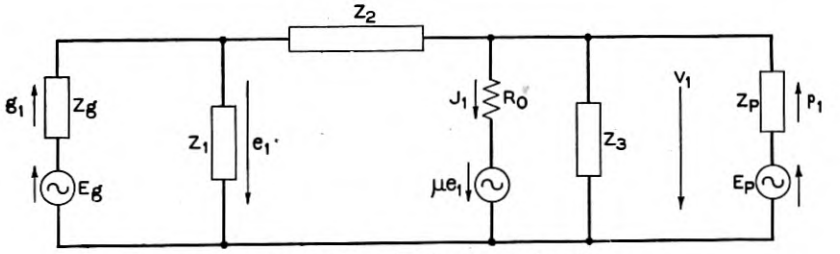


Fig. 2—Equivalent circuit, first order effects.

The equivalence expressed by Fig. 2 is the familiar circuit which has found such wide application, for instance, in amplifier and oscillator work; while the equivalent circuits in Figs. 3 and 4 represent the second and third order effects. With no feedback, that is when Z_2 is infinite, they reduce to the equivalences given by Carson.[1] Comparing now any two equivalent circuits for same order effects with and without feedback we find different values of the electromotive forces appearing in series with the internal tube resistance R_0. Otherwise

[1] Loc. cit., equations (23) and following.

the two circuits are identical except that for one the impedance Z_2 is finite and for the other infinite.

By the aid of the equivalent circuits given, that is by using equations (10), (11), (12), (13) and so forth, the terms in the series (4) can be calculated. These series formally satisfy equations (2) and (3) and are the solutions if they converge.

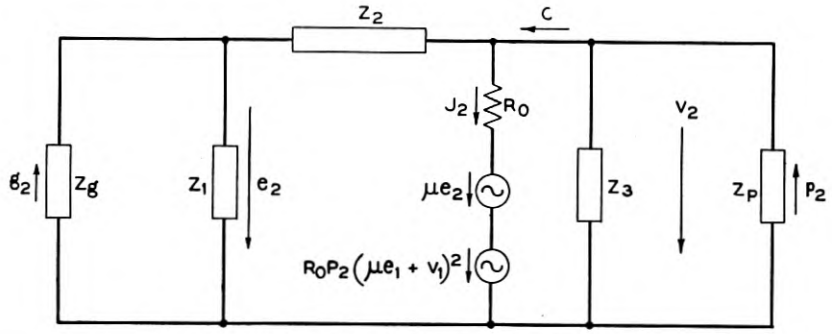


Fig. 3—Equivalent circuit, second order effects.

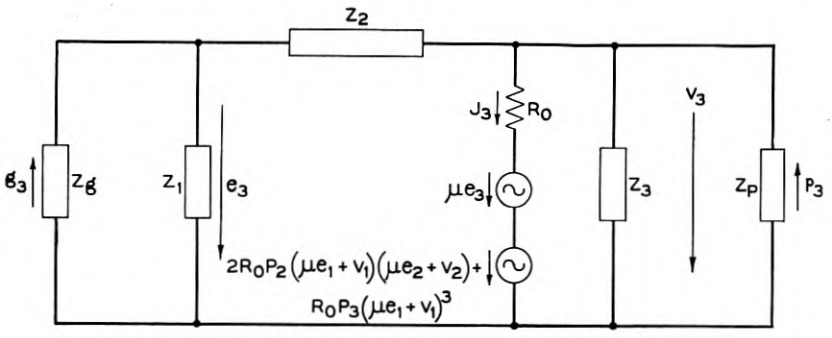


Fig. 4—Equivalent circuit, third order effects.

For the purpose of fixing our ideas we assumed at the start a definite circuit to which the tube was connected. It is obvious, however, that no matter how complicated the linear network is to which the input and output terminals of the tube are connected the procedure given above can be followed.

Application to a Modulator Circuit

As an illustration of the theory just presented we shall calculate the steady state second order effect assuming the circuit configuration to be that given in Fig. 1. In so doing we shall assume that no variable

e.m.f. is impressed in the plate circuit and that the impressed e.m.f. in the grid circuit is given by [6]

$$e_g = K \cos \omega_1 t + S \cos \omega_2 t. \tag{14}$$

We now find the instantaneous value of $\mu e_1 + v_1$ by solving the mesh equations for the equivalent circuit of Fig. 2. The result is:

$$\mu e_1 + v_1 = R_0 \left[\left| \frac{F(\omega_1)}{Z(\omega_1)} \right| K \cos (\omega_1 t - \varphi(\omega_1)) + \left| \frac{F(\omega_2)}{Z(\omega_2)} \right| S \cos (\omega_2 t - \varphi(\omega_2)) \right], \tag{15}$$

where

$$\begin{aligned} F(\omega) &= \frac{Z_1(\mu Z_2 + \mu Z_p' + Z_p')}{(Z_g + Z_1)(Z_g' + Z_2)}, \\ Z(\omega) &= R_0 + Z_p' + \frac{Z_p'(R_0 + \mu Z_g')}{Z_2 + Z_g'}, \\ Z_p' &= \frac{Z_3 Z_p}{Z_3 + Z_p}, \quad Z_g' = \frac{Z_1 Z_g}{Z_1 + Z_g}, \\ \frac{F(\omega)}{Z(\omega)} &= \left| \frac{F(\omega)}{Z(\omega)} \right| e^{-\varphi(\omega) i} (i = \sqrt{-1}). \end{aligned} \tag{16}$$

In equations (16) we note that Z_1, Z_2, Z_3, Z_g and Z_p all are complex impedances. The driving e.m.f. for the second approximation is $R_0 P_2(\mu e_1 + v_1)^2$. Letting

$$M = R_0^3 P_2, \tag{17}$$

we get from (15)

$$\begin{aligned} &R_0 P_2(\mu e_1 + v_1)^2 \\ &= M \left[\frac{1}{2} \left(\left| \frac{F(\omega_1)}{Z(\omega_1)} \right|^2 K^2 + \left| \frac{F(\omega_2)}{Z(\omega_2)} \right|^2 S^2 \right) \right. \\ &\quad + \frac{1}{2} \left| \frac{F(\omega_1)}{Z(\omega_1)} \right|^2 K^2 \cos (2\omega_1 t - 2\varphi(\omega_1)) \\ &\quad + \frac{1}{2} \left| \frac{F(\omega_2)}{Z(\omega_2)} \right|^2 S^2 \cos (2\omega_2 t - 2\varphi(\omega_2)) \\ &\quad + \left| \frac{F(\omega_1) F(\omega_2)}{Z(\omega_1) Z(\omega_2)} \right| KS \cos ((\omega_1 - \omega_2)t - \varphi(\omega_1) + \varphi(\omega_2)) \\ &\quad \left. + \left| \frac{F(\omega_1) F(\omega_2)}{Z(\omega_1) Z(\omega_2)} \right| KS \cos ((\omega_1 + \omega_2)t - \varphi(\omega_1) - \varphi(\omega_2)) \right]. \end{aligned} \tag{18}$$

[6] The extension to any number of sinusoidal e.m.f.'s of arbitrary phases in both plate and grid circuit is obvious.

The driving e.m.f. given by (18) thus consists of a number of sinusoidal components including one of zero frequency. By means of the superposition theorem and the mesh equations we obtain the current and voltage distribution for our equivalent circuit in Fig. 3. Let us for instance calculate the instantaneous current flowing through the impedance consisting of Z_p and Z_3 in parallel and indicated by C in Fig. 3. The result is

$$\begin{aligned} C = M\Bigg[& \frac{\frac{1}{2}\left(\left|\frac{F(\omega_1)}{Z(\omega_1)}\right|^2 K^2 + \left|\frac{F(\omega_2)}{Z(\omega_2)}\right|^2 S^2\right)}{Z(0)} + \\ & + \frac{\frac{1}{2}\left|\frac{F(\omega_1)}{Z(\omega_1)}\right|^2 K^2}{|Z(2\omega_1)|}\cos\left(2\omega_1 t - 2\varphi(\omega_1) - \psi(2\omega_1)\right) \\ & + \frac{\frac{1}{2}\left|\frac{F(\omega_2)}{Z(\omega_2)}\right|^2 S^2}{|Z(2\omega_2)|}\cos\left(2\omega_2 t - 2\varphi(\omega_2) - \psi(2\omega_2)\right) \\ & + \frac{\left|\frac{F(\omega_1)F(\omega_2)}{Z(\omega_1)Z(\omega_2)}\right| KS}{|Z(\omega_1 - \omega_2)|}\cos\left((\omega_1 - \omega_2)t - \varphi(\omega_1)\right. \\ & \qquad \left. + \varphi(\omega_2) - \psi(\omega_1 - \omega_2)\right) \\ & + \frac{\left|\frac{F(\omega_1)F(\omega_2)}{Z(\omega_1)Z(\omega_2)}\right| KS}{|Z(\omega_1 + \omega_2)|}\cos\left((\omega_1 + \omega_2)t - \varphi(\omega_1)\right. \\ & \qquad \left. - \varphi(\omega_2) - \psi(\omega_1 + \omega_2)\right)\Bigg], \end{aligned} \tag{19}$$

where $\psi(\omega)$ is defined by

$$Z(\omega) = |Z(\omega)|e^{i\psi(\omega)} (i = \sqrt{-1}). \tag{20}$$

Let us now consider the peak value of the current of lower side-band frequency. This value is from (19)

$$MKS\left|\frac{F(\omega_1)F(\omega_2)}{Z(\omega_1)Z(\omega_2)Z(\omega_1 - \omega_2)}\right|. \tag{21}$$

If we write

$$|Z(\omega)| = \left|\frac{R_0}{1 + \frac{R_0 + \mu Z_g'}{Z_2 + Z_g'}} + Z_p'\right|_\omega \left|1 + \frac{R_0 + \mu Z_g'}{Z_2 + Z_g'}\right|_\omega, \tag{22}$$

$$|F(\omega)| = (\mu + 1)\left|\frac{Z_1}{(Z_g + Z_1)(Z_g' + Z_2)}\right|_\omega \left|\frac{\mu}{\mu + 1}Z_2 + Z_p'\right|_\omega, \tag{23}$$

expression (21) becomes:

$$\frac{MKS(\mu+1)^2\left|\frac{Z_1}{(Z_g+Z_1)(Z_g'+Z_2)}\right|_{\omega_1}\times\left|\frac{Z_1}{(Z_g+Z_1)(Z_g'+Z_2)}\right|_{\omega_2}\times\left|\frac{\mu}{\mu+1}Z_2+Z_p'\right|_{\omega_1}\left|\frac{\mu}{\mu+1}Z_2+Z_p'\right|_{\omega_2}}{\left|1+\frac{R_0+\mu Z_g'}{Z_2+Z_g'}\right|_{\omega_1}\left|1+\frac{R_0+\mu Z_g'}{Z_2+Z_g'}\right|_{\omega_2}\left|1+\frac{R_0+\mu Z_g'}{Z_2+Z_g'}\right|_{(\omega_1-\omega_2)}}\times\left|Z_p'+\frac{R_0}{1+\frac{R_0+\mu Z_g'}{Z_2+Z_g'}}\right|_{\omega_1}\left|Z_p'+\frac{R_0}{1+\frac{R_0+\mu Z_g'}{Z_2+Z_g'}}\right|_{\omega_2}\times\left|Z_p'+\frac{R_0}{1+\frac{R_0+\mu Z_g'}{Z_2+Z_g'}}\right|_{(\omega_1-\omega_2)}. \quad (24)$$

With no feedback present, that is when $Z_2 = \infty$, equation (24) reduces to

$$\frac{MKS\mu^2\left|\frac{Z_1}{Z_1+Z_g}\right|_{\omega_1}\left|\frac{Z_1}{Z_1+Z_g}\right|_{\omega_2}}{|Z_p'+R_0|_{\omega_1}|Z_p'+R_0|_{\omega_2}|Z_p'+R_0|_{(\omega_1-\omega_2)}}. \quad (25)$$

But in this case $K\left|\frac{Z_1}{Z_1+Z_g}\right|_{\omega_1}$ is the peak value K' of the grid voltage of frequency ω_1 and similarly $S\left|\frac{Z_1}{Z_1+Z_2}\right|_{\omega_2}$ is the peak value S' of the grid voltage of frequency ω_2. Expression (25) may thus be written:

$$\frac{M\mu^2K'S'}{|Z_p'+R_0|_{\omega_1}|Z_p'+R_0|_{\omega_2}|Z_p'+R_0|_{(\omega_1-\omega_2)}},$$

which is the well known expression given by Carson.[7]

For the purpose of getting an idea of the magnitudes involved let us consider a numerical example. A Western Electric No. 101D vacuum tube may have the following constants when used as a modulator: $R_0 = 9000$ ohms; $\mu = 6$; grid-cathode capacitance $C_1 = 10.5$, plate-grid capacitance $C_2 = 4.8$, and plate-cathode capacitance $C_3 = 8.1$ micromicrofarads. The impedances Z_p and Z_g are assumed to be pure resistances at all frequencies with the values 9000 and 10,000 ohms, respectively. The impressed e.m.f. is of the form given by

[7] Carson: *I. R. E. Proc.*, June, 1921, page 243.

the equation (14) where $\omega_1/2\pi$ is equal to 250,000 and $\omega_2/2\pi$ is equal to 5,000,000 cycles per second. As a reference condition let us take that for which the effect of the interelectrode capacitances is neglected. The plate current for this condition is obtained from equation (25) when Z_1 and Z_3 as well as Z_2 are made infinite. As a next step we compute the plate current from equation (24) when Z_1, Z_2, and Z_3 are the impedances corresponding to the interelectrode capacitances C_1, C_2, and C_3, respectively. It is found that this plate current is 12 db below that obtained in the reference condition. Finally it is of some interest to compute the plate current when the grid-plate capacitance alone is effective. This plate current is obtained from equation (24) by assuming Z_1 and Z_3 to be infinite and Z_2 to be the impedance corresponding to the capacitance C_2. This current is found to be 24 db below that of the reference condition.

The Electromagnetic Theory of Coaxial Transmission Lines and Cylindrical Shields

By S. A. SCHELKUNOFF

A form of circuit which is of considerable interest for the transmission of high frequency currents is one consisting of a cylindrical conducting tube within which a smaller conductor is coaxially placed. Such tubes have found application in radio stations to connect transmitting and receiving apparatus to antennæ. As a part of the development work on such coaxial systems, it has been necessary to formulate the theory of transmission over a coaxial circuit and of the shielding against inductive effects which is afforded by the outer conductor. This paper deals generally with the transmission theory of coaxial circuits and extends the theory beyond the range of present application both as regards structure and frequency.

THE mathematical theory of wave propagation along a conductor with an external coaxial return is very old, going back to the work of Rayleigh, Heaviside and J. J. Thomson. Much important work has been done in developing and extending this theory. Among the problems dealt with in this development may be listed the following: the extension of the theory to systems consisting of a plurality of cylindrical conductors; the investigation of shielding and crosstalk in coaxial systems and the effects of eccentricity; the extension of the particular solution to include the complementary modes of propagation, etc.; and in general the adaptation of the mathematical theory to engineering uses, and its translation into the concepts and language of electric circuit theory. In addition to the author's contribution a substantial part of this mathematical work has been done by the group of engineers associated with Mr. John R. Carson, formerly of the American Telephone and Telegraph Company, now of the Bell Telephone Laboratories, Inc.

The problem is ideally adapted to mathematical investigation, because the conductor shape fits perfectly into the cylindrical system of coordinates, thereby making it entirely feasible to carry out a rigorous discussion on the basis of the electromagnetic theory, instead of using ordinary circuit theory. This has obvious advantages at ultra high frequencies, where the uncertainties of the circuit theory are conspicuous and not easily compensated for. It also proves to be of greater advantage at lower frequencies than one might at first assume. Fortunately, it turns out that the final results obtained by means of field theory can be expressed in a familiar language of circuit

theory, thereby gaining all the simplicity of the latter combined with all the accuracy of the former.

Circularly Symmetric Electromagnetic Fields

In polar coordinates, Maxwell's equations assume the following form:

$$\begin{aligned}
&\frac{\partial H_z}{\rho\partial\varphi} - \frac{\partial H_\varphi}{\partial z} = (g + i\omega\epsilon)E_\rho, \qquad \frac{\partial E_z}{\rho\partial\varphi} - \frac{\partial E_\varphi}{\partial z} = -i\omega\mu H_\rho,\\
&\frac{\partial H_\rho}{\partial z} - \frac{\partial H_z}{\partial\rho} = (g + i\omega\epsilon)E_\varphi, \qquad \frac{\partial E_\rho}{\partial z} - \frac{\partial E_z}{\partial\rho} = -i\omega\mu H_\varphi,\\
&\frac{1}{\rho}\left[\frac{\partial(\rho H_\varphi)}{\partial\rho} - \frac{\partial H_\rho}{\partial\varphi}\right] = (g + i\omega\epsilon)E_z,\\
&\frac{1}{\rho}\left[\frac{\partial(\rho E_\varphi)}{\partial\rho} - \frac{\partial E_\rho}{\partial\varphi}\right] = -i\omega\mu H_z,
\end{aligned} \tag{1}$$

where E and H are respectively the electromotive and magnetomotive intensities.[1]

In general, all six field components depend upon each other. If, however, these quantities are independent of either φ or z, the partial derivatives with respect to the corresponding variable vanish and the original set of equations breaks up into two independent subsets, each involving only three physical quantities. Each of these special fields has important practical applications.

In the *circularly symmetric* case, that is, when the quantities are independent of φ, one of the independent subsets is composed of the first and the third equations on the left of (1), together with the second on the right:

$$\begin{aligned}
&\frac{\partial(\rho H_\varphi)}{\partial\rho} = (g + i\omega\epsilon)\rho E_z, \qquad \frac{\partial H_\varphi}{\partial z} = -(g + i\omega\epsilon)E_\rho,\\
&\frac{\partial E_z}{\partial\rho} - \frac{\partial E_\rho}{\partial z} = i\omega\mu H_\varphi.
\end{aligned} \tag{2}$$

This *circular magnetic* field, with its lines of magnetomotive intensity

[1] In this paper we have adopted a unified practical system of units based upon the customary cgs system augmented by adding one typically electric unit. This system has three obvious advantages: first, theoretical results are expressed directly in the units habitually employed in the laboratory; second, the dimensional character and physical significance of such quantities as $i\omega\mu$ and $g + i\omega\epsilon$ are not obscured as in other systems by suppressing dimensions of some electrical unit such as permeability or dielectric constant; and third, the form of electromagnetic equations is very simple. In this system of units the electromotive intensity E is measured in volts/cm., the magnetomotive intensity H in amperes/cm., the intrinsic conductance g in mhos/cm., the intrinsic inductance μ in henries/cm., and the intrinsic capacity ϵ in farads/cm. Thus, in empty space $\mu = 4\pi 10^{-9}$ henries/cm. or approximately 0.01257 μh/cm. and $\epsilon = (1/36\pi)\cdot 10^{-11}$ farads/cm. or approximately 0.0884 mmf./cm.

forming a system of coaxial circles, is associated with currents flowing in isolated wires as, for example, in a single vertical antenna and under ordinary operating conditions it is also found between the conductors of a coaxial pair (Fig. 1).

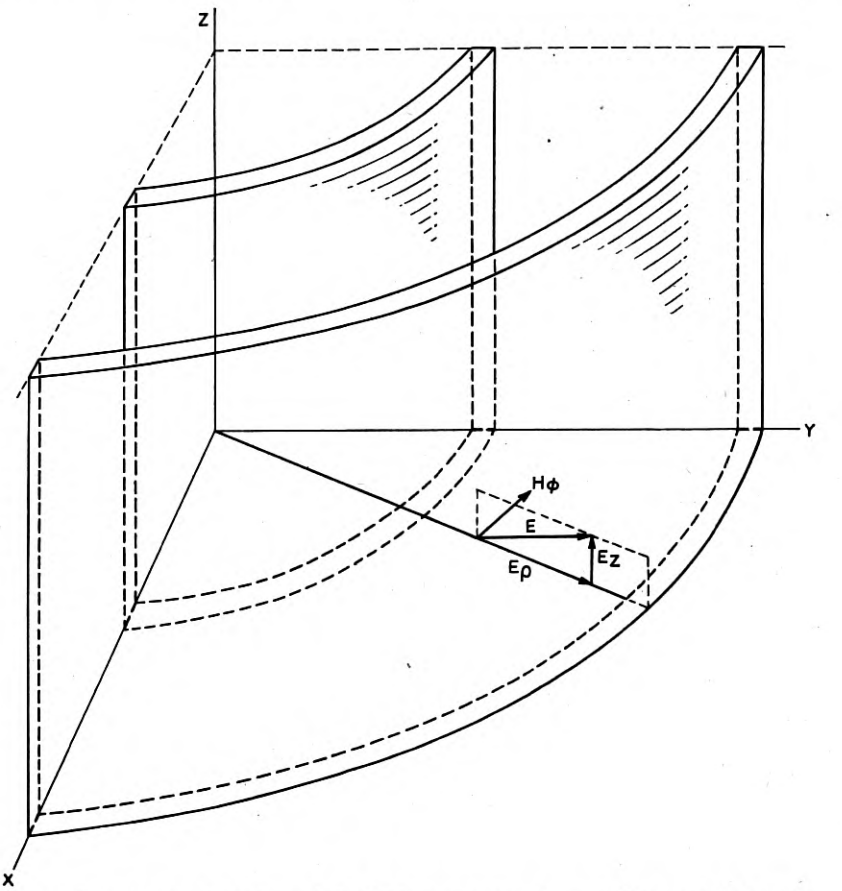


Fig. 1—The relative directions of the field components in a coaxial transmission line.

The remaining three equations of the set (1) form the second group:

$$\frac{\partial(\rho E_\varphi)}{\partial \rho} = -i\omega\mu\rho H_z, \qquad \frac{\partial E_\varphi}{\partial z} = i\omega\mu H_\rho,$$
$$(g + i\omega\epsilon)E_\varphi = \frac{\partial H_\rho}{\partial z} - \frac{\partial H_z}{\partial \rho}, \tag{3}$$

describing the *circular electric* field. Uniformly distributed electric

current in a circular turn of wire is surrounded by a field of this type; in this case, the lines of electromotive intensity form a coaxial system of circles.

Two-Dimensional Fields

By definition, two-dimensional fields are constant in some one direction. If we take the z-axis of our reference system in this direction, all the partial derivatives with respect to z vanish, z disappears from our equations and we can confine our attention to any plane normal to the z-axis.

Once more the set of six electromagnetic equations breaks up into two independent subsets. One of these [2]

$$E_\rho = -\frac{1}{(g + i\omega\epsilon)\rho}\frac{\partial H_z}{\partial \varphi}, \qquad E_\varphi = -\frac{1}{g + i\omega\epsilon}\frac{\partial H_z}{\partial \rho},$$
$$\frac{1}{\rho}\left[\frac{\partial(\rho E_\varphi)}{\partial \rho} + \frac{\partial E_\rho}{\partial \varphi}\right] = -i\omega\mu H_z. \tag{4}$$

The calculation of what is commonly known as "electrostatic" crosstalk between pairs of parallel wires is based upon these equations. For this reason we shall name the field defined by (4) the *electric field.*

Similarly, the remaining three equations define the magnetic field:

$$H_\rho = -\frac{1}{i\omega\mu\rho}\frac{\partial E_z}{\partial \varphi}, \qquad H_\varphi = -\frac{1}{i\omega\mu}\frac{\partial E_z}{\partial \rho},$$
$$\frac{1}{\rho}\left[\frac{\partial(\rho H_\varphi)}{\partial \rho} + \frac{\partial H_\rho}{\partial \varphi}\right] = -(g + i\omega\epsilon)E_z \tag{5}$$

and are useful in the theory of what is generally known as "electromagnetic" crosstalk.

The distinction between electric and magnetic fields is purely pragmatic and is based upon a necessary and valid engineering separation of general electromagnetic interference into two component parts. In some respects the firmly entrenched terms "electrostatic crosstalk" and "electromagnetic crosstalk" are unfortunate; it would be hopeless, however, to try a change of terminology at this late stage of engineering development.

Further consideration of two-dimensional fields will be deferred until the problem of shielding is taken up later in this paper (page 567).

[2] In passing from the original set (1) we reversed the sign of E_ρ in order to make the set of equations symmetrical. The positive E_ρ is now measured toward the axis.

[3] In these equations, the sign of H_φ was reversed so that the magnetomotive intensity is now positive when it points clockwise. With this convention, the flow of energy is away from the axis when both H_φ and E_z are positive.

Exponential Propagation

While electromotive forces could be applied in such a way that the fields would be of the kind given by (3), in the coaxial transmission line as actually energized the fields are of the circular magnetic type (2) which will claim our special attention in the next few sections.

In order to solve equations (2), we naturally want to eliminate all variables but one. This purpose can be readily accomplished if E_z and E_ρ are substituted from the first and the last equations of the set into the second. Thus, we obtain the following equation for the magnetomotive intensity:

$$\frac{\partial}{\partial \rho}\left[\frac{1}{\rho}\frac{\partial(\rho H_\varphi)}{\partial \rho}\right] + \frac{\partial^2 H_\varphi}{\partial z^2} = \sigma^2 H_\varphi, \tag{6}$$

where

$$\sigma^2 = g\omega\mu i - \omega^2\epsilon\mu. \tag{7}$$

Adopting the usual method of searching for particular solutions of (6) in the form

$$H_\varphi = R(\rho)Z(z), \tag{8}$$

where $R(\rho)$ is a function of ρ alone, and $Z(z)$ a function of z alone, we get

$$\frac{1}{Z}\frac{d^2Z}{dz^2} = \Gamma^2, \tag{9}$$

$$\frac{1}{R}\frac{d}{d\rho}\left[\frac{1}{\rho}\frac{d(\rho R)}{d\rho}\right] = \sigma^2 - \Gamma^2, \tag{10}$$

where Γ is some constant about which we have no information for the time being.

Equation (9) is well known in transmission line theory; its general solution can be written in the form

$$Z = Ae^{\Gamma z} + Be^{-\Gamma z}, \tag{11}$$

where A and B are arbitrary constants. The solutions of (10) are Bessel functions. Since equation (6) is linear, we may invoke the principle of superposition and add any number of particular solutions corresponding to different values of Γ. Thus we can form an infinite variety of other solutions so as to satisfy the physical conditions of various practical problems.

It is seen at once from the first and the last equations of the set (2) that to each H_φ of the form (8) there correspond an E_z and E_ρ of the same form; i.e., there exist circularly symmetric electromagnetic

fields, all of whose components vary exponentially in the direction of the axis of symmetry. Whether any of these fields can be produced individually by some simple physical means is impossible to decide on theoretical grounds alone. It may happen, of course, that the field due to *any* practically realizable source is always a combination of several simple exponential fields. In any case, however, we want to know the properties of pure exponential solutions.

It is convenient to make the exponential character of the quantities E_ρ, E_z and H_φ explicit and write them respectively in the form $E_\rho e^{-\Gamma z}$, $E_z e^{-\Gamma z}$ and $H_\varphi e^{-\Gamma z}$. The new quantities E_ρ, E_z and H_φ are functions of ρ only. If the suggested substitution is made in equations (2), the factor $e^{-\Gamma z}$ cancels out and we have

$$E_\rho = \frac{\Gamma}{g + i\omega\epsilon} H_\varphi, \qquad i\omega\mu H_\varphi = \frac{dE_z}{d\rho} + \Gamma E_\rho, \qquad \frac{d(\rho H_\varphi)}{d\rho} = (g + i\omega\epsilon)\rho E_z. \tag{12}$$

The quantity Γ is called the *longitudinal propagation constant* or simply the *propagation constant* when no confusion is possible.[4]

Recalling the implied exponential time factor $e^{i\omega t}$, we see that the complete exponential factor in the expressions for the field intensities is $e^{-\Gamma z + i\omega t}$. The propagation constant Γ is often a complex number and can be represented in the form $\alpha + i\beta$ where the real part is called the *attenuation constant* and the imaginary part, the *phase constant*. Thus, $e^{-\alpha z}$ measures the decrease in the amplitudes of the intensities and $e^{-i(\beta z - \omega t)}$, the change of their phases in time as well as in the z-direction. The latter factor suggests that we are dealing with a wave moving in the positive direction of the z-axis with a velocity (ω/β). A wave moving in the opposite direction is obtained by reversing the sign of Γ.

Perfectly Conducting Coaxial Cylinders [5]

Let us now consider one of the simplest problems which, though purely academic in itself, will throw some light on what is likely to happen under less ideal conditions. We suppose that a perfect dielectric is enclosed between two perfectly conducting coaxial cylinders (Fig. 1) whose radii [6] are b and a $(b < a)$. Our problem is to find the symmetric electromagnetic fields which can exist in such a medium.

[4] Another set of exponential solutions is obtained from this by changing Γ into $-\Gamma$.

[5] For a thorough discussion of "complementary" waves in coaxial pairs the reader is referred to John R. Carson [4].

[6] Only the outer radius of the inner conductor and the inner radius of the outer conductor need be considered because in perfectly conducting media electric states are entirely surface phenomena.

In a perfect dielectric $g = 0$ and the preceding set of equations becomes

$$E_\rho = \frac{\Gamma}{i\omega\epsilon} H_\varphi, \qquad i\omega\mu H_\varphi = \frac{dE_z}{d\rho} + \Gamma E_\rho, \qquad \frac{d(\rho H_\varphi)}{d\rho} = i\omega\epsilon\rho E_z. \tag{13}$$

No force is required to sustain electric current in perfect conductors and the tangential components of the intensities are continuous across the boundaries between different media; therefore, the longitudinal electromotive intensity vanishes where ρ equals either a or b.

Substituting E_ρ from the first equation into the second, solving the latter for H_φ and inserting it into the third equation, we have successively

$$H_\varphi = -\frac{i\omega\epsilon}{m^2}\frac{dE_z}{d\rho}, \tag{14}$$

and

$$\rho\frac{d^2E_z}{d\rho^2} + \frac{dE_z}{d\rho} + m^2\rho E_z = 0, \tag{15}$$

where, for convenience, we let $\Gamma^2 + \omega^2\epsilon\mu = m^2$. The most general solution of the last equation is usually written in the form

$$E_z(\rho) = AJ_0(m\rho) + BY_0(m\rho), \tag{16}$$

where J_0 and Y_0 are Bessel functions of order zero and A and B are constants so far unknown.[7]

The constants A and B can be determined from the fact already mentioned that E_z vanishes on the surface of either conductor, i.e., from the following equations:

$$AJ_0(mb) + BY_0(mb) = 0, \qquad \text{and} \qquad AJ_0(ma) + BY_0(ma) = 0. \tag{17}$$

These equations are certainly satisfied if both constants are equal to 0. If, however, they are not equal to 0 simultaneously, we can determine their ratio from each equation of the above system. These ratios should be the same, of course, and yet they cannot be equal for every value of m. Thus, the *permissible* values of m are the roots of

[7] For large values of the argument these Bessel functions are very much like slightly damped sinusoidal functions; in fact $J_0(x)$ and $Y_0(x)$ are approximately equal, respectively, to $\sqrt{2/\pi x}\cos(x - \pi/4)$ and $\sqrt{2/\pi x}\sin(x - \pi/4)$, provided x is large enough.

the following equation:

$$-\frac{A}{B} = \frac{Y_0(mb)}{J_0(mb)} = \frac{Y_0(ma)}{J_0(ma)} . \tag{18}$$

This equation has an infinite number of roots [8] whose approximate values can be readily determined if we replace Bessel functions by their approximations in terms of circular functions. Thus, we have

$$m_n = \frac{n\pi}{a - b} , \qquad (n = 1, 2, 3, \cdots). \tag{19}$$

This is a surprisingly good approximation for *all* roots if the radius of the outer conductor is less than three times that of the inner; and the larger the n, the better the approximation.[9] The propagation constants are computed from the corresponding values of m_n by means of the following equation,

$$\Gamma_n = \sqrt{m_n^2 - \omega^2\epsilon\mu}. \tag{20}$$

First of all, let us study the simplest solution in which both A and B vanish. In this case, the longitudinal electromotive intensity vanishes identically. The magnetomotive intensity—and the transverse electromotive intensity, as well—also vanishes unless the denominator m^2 in equation (14) equals zero. If all intensities were to vanish, we should have no field and there would be nothing to talk about; hence, we take the other alternative and let

$$\Gamma^2 + \omega^2\epsilon\mu = 0, \qquad \text{i.e.,} \qquad \Gamma = i\omega\sqrt{\epsilon\mu}, \tag{21}$$

the positive sign having been implied in writing equations (13). In air, $\epsilon\mu = (1/c^2)$ where c is the velocity of light in cm.; hence, in air this particular propagation constant equals $i\omega/c$. Since E_z equals zero everywhere, the electromotive intensity is wholly transverse; and the flow of energy being, according to Poynting, at right angles to the electromotive and magnetomotive intensities, the energy transfer is wholly longitudinal.

The above method of determining the propagation constant may be open to suspicion; besides, the method does not tell how to obtain the actual values of the electromagnetic intensities but merely leads to a relation compatible with the existence of such intensities. Therefore, let us obtain the wanted information directly from the funda-

[8] A. Gray and G. B. Mathews, "A Treatise on Bessel Functions" (1922), p. 261.

[9] It is strictly accurate if the radii of the cylinders are infinite, i.e., if we are dealing with a dielectric slab bounded by perfectly conducting planes.

mental equations (13) which assume the following simple form:

$$i\omega\epsilon E_\rho = \Gamma H_\varphi, \qquad i\omega\mu H_\varphi = \Gamma E_\rho, \qquad \frac{d(\rho H_\varphi)}{d\rho} = 0, \tag{22}$$

if E_z vanishes identically. Either of the first two equations determines the ratio of the electromotive intensity to the magnetomotive; the two ratios are consistent only if the condition (21) is satisfied. Then, we have also

$$E_\rho = \frac{\Gamma}{i\omega\epsilon} H_\varphi = \sqrt{\frac{\mu}{\epsilon}} H_\varphi \qquad \text{and} \qquad H_\varphi = \frac{A}{\rho}, \tag{23}$$

where A is some quantity independent of ρ. This constant can be readily calculated from Ampere's law. The magnetomotive force acting along the circumference of any particular cross-section of the inner cylinder equals $2\pi\rho H_\varphi$ amperes, i.e., $2\pi A$; since this M.M.F. should equal the total current I flowing in the inner conductor through the cross-section, the quantity A equals $I/2\pi$. Reintroducing the implied factor $e^{-\Gamma z}$, we have

$$\begin{aligned} H_\varphi &= \frac{I}{2\pi\rho} e^{-\Gamma z}, \\ E_\rho &= \frac{I}{2\pi\rho} \sqrt{\frac{\mu}{\epsilon}} e^{-\Gamma z}. \end{aligned} \tag{24}$$

In practical measurements we are concerned with the total potential difference (V) between the cylinders, rather than with the transverse electromotive intensity. The former is merely the integral of the intensity,

$$V = \int_b^a E_\rho d\rho = \left(\frac{1}{2\pi} \sqrt{\frac{\mu}{\epsilon}} \log \frac{a}{b} \right) I e^{-\Gamma z}. \tag{25}$$

This voltage and the current I vary as voltage and current in a semi-infinite transmission line whose propagation constant is Γ *and whose characteristic impedance is*

$$Z_0 = \frac{V}{I e^{-\Gamma z}} = \frac{1}{2\pi} \sqrt{\frac{\mu}{\epsilon}} \log \frac{a}{b}. \tag{26}$$

At any point z the intensities E_ρ *and* H_φ *have the same values as would the voltage and current at the same distance z from the end of a transmission line whose propagation constant and characteristic impedance are respectively* $i\omega\sqrt{\epsilon\mu}$ *and* $\sqrt{\mu/\epsilon}$.

The connection between electromagnetic theory and line theory is so important that, risking repetition, we wish to emphasize their intimate relationship by deriving the well-known differential equations of the line theory directly from the electromagnetic equations (2) *combined with the assumption that the longitudinal electromotive intensity vanishes everywhere.* We already know that under the assumed conditions the first equation of the system (2) becomes

$$H_\varphi = \frac{I}{2\pi\rho}, \tag{27}$$

where I is the total current flowing in the inner cylinder through a particular cross-section and is some function [10] of z. We can therefore rewrite the last two equations of the system as follows:

$$\frac{\partial E_\rho}{\partial z} = -\frac{i\omega\mu}{2\pi\rho} I, \qquad \frac{1}{2\pi\rho}\frac{\partial I}{\partial z} = -i\omega\epsilon E_\rho. \tag{28}$$

We have merely to integrate both equations with respect to ρ from b to a and substitute the potential difference V for the integral of the transverse electromotive intensity to obtain

$$\frac{\partial V}{\partial z} = -\left(\frac{i\omega\mu}{2\pi}\log\frac{a}{b}\right) I, \qquad \frac{\partial I}{\partial z} = -\frac{2\pi i\omega\epsilon}{\log\frac{a}{b}} V, \tag{29}$$

which are the equations of the transmission line whose distributed series inductance equals $(\mu/2\pi)\log(a/b)$ henries/cm. and shunt capacity $2\pi\epsilon/(\log a/b)$ farads/cm.

With this, we conclude the special case in which the longitudinal electromotive intensity vanishes everywhere, the propagation constant equals $i\omega\sqrt{\epsilon\mu}$, and the velocity of transmission is that of light.

We now turn our attention to the case in which A and B do not vanish. We have already noted that the propagation constants are given by equation (20). Since, in this case, we are interested primarily in the nature of the phenomena rather than in the details of field distribution, we shall simplify our mathematics by supposing the radii of the cylinders to be infinite. Thus, the cylinders become two planes perpendicular to the x-axis, distance a apart. The φ-direction, then, coincides with the y-direction and, therefore, all the intensities are independent of the y-coordinate. Let us choose the z-axis halfway between the planes. The equations describing this two-dimen-

[10] On this occasion, we should remember that a particular type of this function had not yet been ascertained at the time the equations (2) were arrived at.

sional transmission line are

$$\frac{\partial H_y}{\partial x} = i\omega\epsilon E_z,$$
$$\frac{\partial H_y}{\partial z} = -i\omega\epsilon E_x, \tag{30}$$
$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -i\omega\mu H_y.$$

If n is an odd integer, these possess the following solutions:

$$E_x = A\frac{\Gamma_n}{i\omega\epsilon}\sin\frac{n\pi x}{a},$$
$$E_z = A\frac{n\pi}{i\omega\epsilon a}\cos\frac{n\pi x}{a}, \tag{31}$$
$$H_y = A\sin\frac{n\pi x}{a};$$

and if n is an even integer,

$$E_x = A\frac{\Gamma_n}{i\omega\epsilon}\cos\frac{n\pi x}{a},$$
$$E_z = -A\frac{n\pi}{i\omega\epsilon a}\sin\frac{n\pi x}{a}, \tag{32}$$
$$H_y = A\cos\frac{n\pi x}{a},$$

where

$$\Gamma_n = \sqrt{\frac{n^2\pi^2}{a^2} - \omega^2\epsilon\mu} = \pi\sqrt{\frac{n^2}{a^2} - \frac{4}{\lambda^2}}, \tag{33}$$

and λ is the wave-length corresponding to the frequency f.

Let us now define the *longitudinal impedance* (Z_z) as the ratio of E_x to H_y,

$$Z_z = \frac{\Gamma_n}{i\omega\epsilon}, \tag{34}$$

and the *transverse impedance* (the impedance in the x-direction) as the ratio of E_z to H_y,

$$Z_x = \frac{n\pi}{i\omega\epsilon a}\cot\frac{n\pi x}{a}, \qquad \text{if } n \text{ is odd},$$
$$Z_x = -\frac{n\pi}{i\omega\epsilon a}\tan\frac{n\pi x}{a}, \qquad \text{if } n \text{ is even}. \tag{35}$$

It will be observed that, depending on the frequency, the longitudinal

propagation constant Γ_n is either real or purely imaginary; it vanishes if $a = n(\lambda/2)$, that is, if the spacing between the planes is a whole number of half wave-lengths. When the propagation constant is real, the longitudinal impedance is purely imaginary, and vice versa, when the propagation constant is purely imaginary, the longitudinal impedance is real. In the former case, no energy is transmitted longitudinally but merely surges back and forth, and in the latter case we have a true transmission line. The transverse impedance is purely imaginary at all frequencies and, hence, the energy merely fluctuates to and fro.

If the frequency is sufficiently low, all of these higher order propagation constants are real and all the energy is transmitted in the *principal mode* described by equations (21) to (29). The rôle of the higher propagation constants consists in redistributing the energy near the sending terminal,[11] that is, in terminal distortion. But as the frequency gets high enough to make the wave-length less than $2a$, the next transmission mode may become prominent, and so forth up the infinite ladder of transmission modes.

Imperfect Coaxial Conductors [12]

We shall now suppose that the conductors are not perfect; i.e., the conductivity instead of being infinite, is merely large. Assuming that our solutions are continuous functions of conductivity (this can be proved), we conclude: first, there exists an infinite series of propagation constants approaching the values given in the preceding section as the conductivity tends to infinity; second, one of these propagation constants, namely that approaching $i\omega\sqrt{\epsilon\mu}$, is very small unless the conductivity is too small. In the immediately succeeding sections we shall be concerned only with electromagnetic fields corresponding to this particular propagation constant.

Let us now prove that the simple expression for the magnetomotive intensity in the dielectric between perfectly conducting cylinders is still true for all practical purposes, even if the conductors are merely good, and even when there are more than two of them. Since the lines of force are circles, coaxial with the conductors, and since H_φ is independent of φ, the total magnetomotive force acting along any one of the circles equals H_φ times the circumference of the circle $(2\pi\rho)$. This M.M.F. also equals the total current I passing through the area of the circle. Therefore, the magnetomotive intensity is $(I/2\pi\rho)$ amperes/cm. This expression is true at any point in the conductors as

[11] And near the receiving terminal as well, if the line is finite.

[12] The general theory of wave propagation in a multiple system of imperfect coaxial conductors is amply covered by John R. Carson and J. J. Gilbert [2, 3].

well as in the dielectric between them. In a conductor the total current I passing through the area of the circle is a function of ρ since the current is distributed throughout the entire cross-section of the conductor. Strictly speaking, the same is true of any circle in the dielectric. There is one important difference, however; the conduction current passing through such circles is the same and the displacement current is usually so small that it can be legitimately neglected. Thus, in the dielectric, we have to an extremely high degree of accuracy unless ρ is very large

$$H_\varphi = \frac{I}{2\pi\rho}, \tag{36}$$

where I is merely a constant, namely, the total *conduction* current passing through the area of the circle of radius ρ.

That the longitudinal displacement current can be neglected, unless the conductivity of the conductors is small, has been already indicated in the opening paragraph. The following comparison is an aid to the mathematical argument. The density of the longitudinal conduction current is gE and that of the displacement current is $i\omega\epsilon E$. Near the boundary, E is substantially the same in the conductor and in the dielectric. In copper, $g = (1/1.724)10^6$ and in air $\epsilon = (1/36\pi)10^{-11}$. Thus, even at very high frequencies, the density of the displacement current is very small compared to that of the conduction current. On the other hand, the conduction current is ordinarily distributed over a small area while the displacement current may flow across a large area. The latter area would have to be very large, however, before it could even begin to compensate for the extremely low current density.

Electromotive Intensities in Dielectrics

With the aid of equations (12) and (36), we can now calculate the electromotive intensities in the dielectric between two conductors. Thus, the transverse intensity is

$$E_\rho = \frac{\Gamma I}{2\pi(g + i\omega\epsilon)\rho}. \tag{37}$$

Substituting this in the second equation of the set (12), we obtain the following differential equation for the longitudinal intensity:

$$\frac{dE_z}{d\rho} = \left[i\omega\mu - \frac{\Gamma^2}{g + i\omega\epsilon} \right] \frac{I}{2\pi\rho}, \tag{38}$$

where μ is the permeability of the dielectric. Integrating with respect

to ρ, we have

$$E_z = \frac{1}{2\pi}\left[i\omega\mu - \frac{\Gamma^2}{g + i\omega\epsilon} \right] I \log \frac{\rho}{b'} + A, \tag{39}$$

where A is a constant to be determined from the boundary conditions.[13]

THE POTENTIAL DIFFERENCE BETWEEN TWO COAXIAL CYLINDERS

Equation (36) relates the transverse electromotive intensity to the total current flowing in the inner conductor. In practice, however, we are interested in the difference of potential between the conductors, that is, in the transverse electromotive force rather than the electromotive intensity. This potential difference V is obtained at once from equation (37) by integration:

$$V = \int_{b'}^{a''} E_\rho d\rho = \frac{\Gamma I}{2\pi(g + i\omega\epsilon)} \int_{b'}^{a''} \frac{d\rho}{\rho} = \frac{\Gamma I \log \frac{a''}{b'}}{2\pi(g + i\omega\epsilon)}. \tag{40}$$

This transverse E.M.F. produces a transverse electric current which is partly a conduction current—if the dielectric is not quite perfect—and partly a displacement (or "capacity") current.

Now, the total transverse current per centimeter length of line is

$$I_\rho = 2\pi\rho(g + i\omega\epsilon)E_\rho.$$

Then, by equation (37), we have

$$I_\rho = \Gamma I. \tag{41}$$

Therefore, equation (40) becomes

$$V = \frac{\log \frac{a''}{b'}}{2\pi(g + i\omega\epsilon)} I_\rho. \tag{42}$$

The ratio of a current to the electromotive force that produces it is called *admittance*. Hence, the distributed *radial admittance* per

[13] The following system of notation will be adhered to throughout the remainder of the paper: The inner radius of any cylindrical conductor is denoted by a, and its outer radius by b. When several coaxial conductors are used, they are differentiated by superscripts; a', a'', $a^{(3)}$, $\cdots$ referring to their inner radii, for example, and b', b'', $b^{(3)}$, $\cdots$ to their outer radii. This convention also applies to conductivities, permeabilities, and other physical constants of the conductors in question.

For convenience, we have written the ratio of ρ to the outer radius of the inner conductor in place of ρ; this change affects only the arbitrary constant A which will eventually be assigned the value required by the boundary conditions. When written in this form, the first term of E_z vanishes on the surface of the inner conductor which is a convenience in determining the value of A.

unit length between two cylindrical conductors is

$$Y = \frac{2\pi(g + i\omega\epsilon)}{\log \frac{a''}{b'}} \equiv G + i\omega C, \tag{43}$$

the symbols G and C being used in the usual way to designate the distributed radial conductance and capacity. Writing these separately, we have

$$G = \frac{2\pi g}{\log \frac{a''}{b'}}, \qquad C = \frac{2\pi\epsilon}{\log \frac{a''}{b'}}. \tag{44}$$

Returning to (40), we find that V can be written in the form

$$V = \frac{\Gamma}{Y} I.$$

But the ratio of the transverse electromotive force V to the longitudinal current I is known as the longitudinal *characteristic impedance* of the coaxial pair. Its value is obviously Γ/Y.

The External Inductance

In dealing with parallel wires it is customary to use the term "external inductance" for the total magnetic flux in the space surrounding the pair.[14] We shall adopt the same usage in connection with coaxial pairs. Strictly speaking, we must therefore consider it as being composed of two parts: one being the flux *between* the cylinders, the other the flux in the space surrounding them. But the longitudinal displacement current is negligible by comparison with the conduction current, and effects due to it have been consistently ignored throughout this part of our study. To the same order of approximation, the flux outside the pair is negligible by comparison with that between them, whence we find the "external inductance" to be

$$L_e = \frac{\mu \int_{b'}^{a''} H_\varphi d\rho}{I} = \frac{\mu}{2\pi} \log \frac{a''}{b'} \text{ henries/cm.} \tag{45}$$

[14] While this definition is very descriptive, it is not strictly accurate unless the wires are perfectly conducting. The correct definition should read as follows: The external inductance of a parallel pair is the measure (per unit current) of magnetic energy stored in the space surrounding the pair. The reason the simpler definition fails for imperfectly conducting parallel wires is because some of the lines of magnetic flux lie partly inside and partly outside the wires. This does not happen in connection with coaxial pairs even when they are not perfectly conducting. Hence we are warranted in using the simpler idea.

Comparing this with equation (44), we have the following relation between the external inductance and the capacity

$$CL_e = \epsilon\mu. \tag{46}$$

Propagation Constants of Coaxial Pairs

Since the relation between electromotive intensity and current is linear, we are justified in writing the intensities at the adjacent surfaces of the pair in the form

$$E_z(b') = Z_b'I, \qquad E_z(a'') = Z_a''I, \tag{47}$$

where Z_b' and Z_a'' depend only upon the material of the conductors and the geometry of the system. These quantities will be called *surface impedances* of the inner and outer conductors, respectively.

Inserting (47) in (39) we obtain

$$\begin{aligned} A &= Z_b'I, \\ \frac{1}{2\pi}\left[i\omega\mu - \frac{\Gamma^2}{g + i\omega\epsilon} \right] I \log\frac{a''}{b'} + A &= - Z_a''I, \end{aligned} \tag{48}$$

by means of which A and Γ may be expressed in terms of Z_b' and Z_a''. If we solve the first of these for A and substitute the value thus derived in the second we get, by virtue of (45),

$$\frac{\Gamma^2}{2\pi(g + i\omega\epsilon)} \log\frac{a''}{b'} = Z_a'' + Z_b'' + i\omega L_e, \tag{49}$$

or, by (43)

$$\Gamma^2 = YZ, \tag{50}$$

where for brevity we have written

$$Z = Z_a'' + Z_b' + i\omega L_e. \tag{51}$$

Direct Conversion of the Circularly Symmetric Field Equations into Transmission Line Equations

As the practical applications of Maxwell's theory become more numerous, it becomes increasingly important to formulate its exact connection with transmission line theory. With this purpose in mind, let us attempt to throw (2) into the form of the transmission line equations.

The obvious plan of attack is to introduce into (2) the transverse voltage V and the longitudinal current I, in place of the intensities E and H. The total current is introduced by substituting $(I/2\pi\rho)$ for H_φ, and the total voltage by integrating the set of equations (2) in the transverse direction. The first equation gives us nothing of

importance.[15] The second and third equations, on the other hand, give

$$\frac{i\omega\mu I}{2\pi}\log\frac{a''}{b'} = E_z''(a) - E_z'(b) - \frac{\partial V}{\partial z}, \tag{52}$$

$$V = -\frac{\log\frac{a''}{b'}}{2\pi(g + i\omega\epsilon)}\frac{\partial I}{\partial z}.$$

But, upon substituting (45), (47) and (51) in the first of these equations and (43) in the second, we get

$$\frac{\partial V}{\partial z} = -ZI, \qquad \frac{\partial I}{\partial z} = -YV, \tag{53}$$

where Z and Y are to be interpreted respectively as the distributed series impedance and shunt admittance.

Current Distribution in Cylindrical Conductors

So far, we have been dealing with electromagnetic intensities in dielectrics. We now turn our attention to conductors and determine their current distributions with the ultimate view of calculating their surface impedances. One of our sources of information is the familiar set of equations (12). In these equations, however, we now let $\epsilon = 0$ since the displacement current in conductors is negligibly small by comparison with the conduction current. From these equations, we eliminate electromotive intensities and thus obtain a differential equation for the magnetomotive intensity. The latter is in fact equation (6) with only one difference: the exponential factor $e^{-\Gamma z}$ has been explicitly introduced and cancelled so that the equation has become

$$\frac{d}{d\rho}\left[\frac{1}{\rho}\frac{d(\rho H_\varphi)}{d\rho}\right] = (\sigma^2 - \Gamma^2)H_\varphi,$$

or

$$\frac{d^2H_\varphi}{d\rho^2} + \frac{1}{\rho}\frac{dH_\varphi}{d\rho} - \frac{H_\varphi}{\rho^2} = (\sigma^2 - \Gamma^2)H_\varphi, \tag{54}$$

where

$$\sigma^2 = g\omega\mu i = 2\pi g\mu f i.$$

This σ will be called the intrinsic propagation constant of solid metal.

[15] Our standard practice of neglecting the longitudinal displacement currents has given us the general rule that $2\pi\rho H_\varphi = I$ is independent of ρ. Using this relation in the first of equations (2.2), we get

$$(g + i\omega\epsilon)E_z \doteq 0;$$

but this merely reflects the fact that $g + i\omega\epsilon$ is very small.

The attenuation and the phase constants are each equal to $\sqrt{\pi g \mu f}$. The intrinsic propagation constants of metals are large quantities except at low frequencies as the accompanying table indicates.

PROPAGATION CONSTANT OF COMMERCIAL COPPER

$g = 5.800\ 10^5$ mhos/cm.
$\mu = 0.01257\ \mu h$/cm.

f	$\frac{\sigma}{1+i} = \sqrt{\pi g \mu f}$
0	0.0
1	0.1513
10	0.4785
100	1.513
10,000	15.13
1,000,000	151.3
100,000,000	1513.

On the other hand, Γ is very small; if air is the dielectric between the conductors, Γ is of the order of $(1/3)i\omega\ 10^{-10}$. Hence, even at high frequencies Γ^2 is negligibly small by comparison with σ^2 and we can rewrite (54) as follows:

$$\frac{d}{d\rho}\left[\frac{1}{\rho}\frac{d}{d\rho}(\rho H_\varphi)\right] = \sigma^2 H_\varphi. \tag{55}$$

This is Bessel's equation and its solution can be written down at once [16] as

$$H_\varphi = AI_1(\sigma\rho) + BK_1(\sigma\rho), \tag{56}$$

where the functions $I_1(u)$ and $K_1(u)$ are the *modified Bessel functions of the first order and respectively of the first and second kind.* For large values of the argument we have approximately

$$\begin{aligned} I_1(u) &= \frac{e^u}{\sqrt{2\pi u}}\left(1 - \frac{3}{8u}\right), \\ K_1(u) &= \sqrt{\frac{\pi}{2u}}\, e^{-u}\left(1 + \frac{3}{8u}\right); \end{aligned} \tag{57}$$

[16] It is interesting to note that in the case of a fairly thin hollow conductor whose inner radius is not too small there exist very simple approximate solutions of (55). Under these circumstances ρ varies over such a small range that no serious error is introduced in treating the factors $(1/\rho)$ and ρ in (55) as constants, and the equation becomes

$$\frac{d^2 H_\varphi}{d\rho^2} = \sigma^2 H_\varphi,$$

which is satisfied by the exponential functions $e^{\sigma\rho}$ and $e^{-\sigma\rho}$. The larger the value of ρ and the faster the change in H_φ with ρ, the better is the approximation.

while for small values

$$I_1(u) = \frac{u}{2} + \frac{u^3}{16},$$
$$K_1(u) = \frac{1}{u} + \frac{u}{2}\log\frac{u}{2}. \tag{58}$$

The function $I_1(u)$ becomes infinite and $K_1(u)$ vanishes when u is infinite.[17] When u is zero, $I_1(u)$ vanishes and $K_1(u)$ becomes infinite.

The longitudinal electromotive intensity is calculated from the third equation (12) with the aid of the following rules for differentiation of modified Bessel functions of any order n:

$$\frac{d}{dx}(x^n I_n) = x^n I_{n-1},$$
$$\frac{d}{dx}(x^n K_n) = -x^n K_{n-1}. \tag{59}$$

Thus,

$$E_z = \eta[AI_0(\sigma\rho) - BK_0(\sigma\rho)], \tag{60}$$

where

$$\eta = \frac{\sigma}{g} = \frac{i\omega\mu}{\sigma}. \tag{61}$$

For reasons which will appear later, this quantity η will be called the *intrinsic impedance of solid metal.*

The current density is merely the product of the intensity E_z and the conductivity g.

In a general way the behavior of the functions of zero order is similar to that of the functions whose order is unity. Thus, for large values of the argument,

$$I_0(u) = \frac{e^u}{\sqrt{2\pi u}}\left(1 + \frac{1}{8u}\right),$$
$$K_0(u) = \sqrt{\frac{\pi}{2u}}e^{-u}\left(1 - \frac{1}{8u}\right), \tag{62}$$

and for small values

$$I_0(u) = 1 - \frac{u^2}{4},$$
$$K_0(u) = -\log u + 0.116. \tag{63}$$

[17] This statement is correct only as long as the real part of u is positive. This is so in our case because out of two possible values of the square root representing σ we can always choose the one with the positive real part.

Surface Impedance of a Solid Wire

On page 547 we defined the surface impedances of a coaxial pair as the ratios of the longitudinal electromotive intensities on the adjacent surfaces of the cylinders to the total currents flowing in the respective conductors. In that place, however, we were unable to give explicit formulæ for the impedances so defined because we did not yet have a precise value for E_z. Now that this omission has been supplied, we are prepared to compute Z_b' and Z_a''.

We consider the case of a *solid* inner cylinder surrounded by *any coaxial return*, and seek to determine the constants A and B in (60). Since the E.M.I. must be finite along the axis of the wire we must make $B = 0$, because the K-function becomes infinite when $\rho = 0$. On the surface of the wire the magnetomotive intensity is $I/2\pi b$ if I is the total current in the wire. By equation (56) this intensity equals $AI_1(\sigma b)$; hence,

$$A = \frac{I}{2\pi b I_1(\sigma b)}$$

and the final expression for the electromotive intensity within the wire is

$$E_z(\rho) = \frac{\eta I_0(\sigma\rho)}{2\pi b I_1(\sigma b)} I. \tag{64}$$

Thus, we have the following expression for the surface impedance of the solid wire:

$$Z_b = \frac{E_z(b)}{I} = \frac{\eta I_0(\sigma b)}{2\pi b I_1(\sigma b)}, \text{ ohms/cm.} \tag{65}$$

As the argument increases, the modified Bessel functions of the first kind (the I-functions) become more and more nearly proportional to the exponential functions of the same argument. Thus, if the absolute value of σb exceeds 50, the Bessel functions in the preceding equation cancel out and the following simple formula holds within 1 per cent:

$$Z_b = \frac{\eta}{2\pi b} = \frac{1}{2b}\sqrt{\frac{\mu f}{\pi g}}(1 + i), \text{ ohms/cm.} \tag{66}$$

This surface impedance consists of a resistance representing the amount of energy dissipated in heat, and a reactance due to the magnetic flux in the wire itself. Separating (66) into these two parts, we have, approximately,

$$R_b = \omega L_b = \frac{1}{2b}\sqrt{\frac{\mu f}{\pi g}}.$$

However, most of the error in (66) occurs in the real part. If more accurate approximations for Bessel functions are used, then

$$\begin{aligned} R_b &= \frac{1}{2b}\sqrt{\frac{\mu f}{\pi g}} + \frac{1}{4\pi g b^2}, \\ \omega L_b &= \frac{1}{2b}\sqrt{\frac{\mu f}{\pi g}}; \end{aligned} \tag{67}$$

these are correct within 1 per cent if $|\sigma b| > 6$. The surface inductance L_b equals $(1/4\pi b)\sqrt{\mu/\pi g f}$ henries/cm.; it decreases as the frequency increases.

If the wire is so thin or the frequency is so low that $|\sigma b| < 6$, equation (65) has to be used. Its use in computations is quite simple, however, because the argument σb is a complex number of the form $u\sqrt{i}$; and the necessary functions have been tabulated. Lord Kelvin introduced the symbols ber u and bei u for the real and the imaginary parts of $I_0(u\sqrt{i})$, so that we now write

$$I_0(u\sqrt{i}) = \text{ber } u + i \text{ bei } u. \tag{68}$$

Differentiating, we have

$$\sqrt{i}\, I_0'(u\sqrt{i}) = \sqrt{i}\, I_1(u\sqrt{i}) = \text{ber}' u + i \text{ bei}' u,$$

and therefore

$$I_1(u\sqrt{i}) = \frac{\text{ber}' u + i \text{ bei}' u}{\sqrt{i}}. \tag{69}$$

If we insert these values in (65), and recall that the d.-c. resistance of a solid wire is $1/\pi g b^2$, and that $\sigma = g\eta$, we obtain at once

$$\begin{aligned} \frac{Z_b}{R_{\text{d.-c.}}} = \pi g b^2 Z_b = \frac{u}{2}\frac{\text{ber } u \text{ bei}' u - \text{bei } u \text{ ber}' u}{(\text{ber}' u)^2 + (\text{bei}' u)^2} \\ + i\frac{u}{2}\frac{\text{ber } u \text{ ber}' u + \text{bei } u \text{ bei}' u}{(\text{ber}' u)^2 + (\text{bei}' u)^2}, \end{aligned} \tag{70}$$

where u is the absolute value of σb. The accompanying graph illustrates the real and imaginary parts of this equation [18] (Fig. 2).

THE SURFACE IMPEDANCES OF HOLLOW CYLINDRICAL SHELLS [19]

In the case of a hollow conductor whose inner and outer radii are respectively equal to a and b, the return coaxial path for the current

[18] For equation (70) and various approximations see E. Jahnke and F. Emde.

[19] In the case of self-impedances the more general equations of two parallel cylindrical shells were deduced by Mrs. S. P. Mead. For the special formulæ concerning self-impedances of coaxial pairs see A. Russell.

may be provided either outside the given conductor or inside it or partly inside and partly outside. We designate by Z_{aa} the surface impedance with internal return, and by Z_{bb}, that with external return. These impedances are equal only at zero frequency; but if the con-

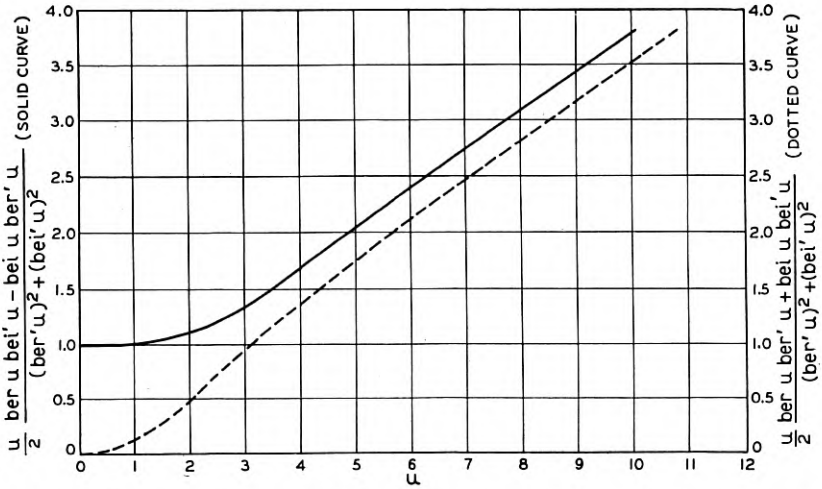


Fig. 2—The skin effect in solid wires. The upper curve represents the ratio of the a-c. resistance of the wire to its d-c. resistance and the lower curve the ratio of the internal reactance to the d-c. resistance.

ductor is thin, they are nearly equal at all frequencies. If the return path is partly internal and partly external, we have in effect two transmission lines with a distributed mutual impedance Z_{ab} due to the mingling of the two currents in the hollow conductor common to both lines. However, since this quantity Z_{ab} is not the total mutual impedance between the two lines unless the hollow conductor is the only part of the electromagnetic field common to them, it is better to call Z_{ab} the *transfer impedance* from one surface of the conductor to the other.

In order to determine these impedances, let us suppose that of the total current $I_a + I_b$ flowing in the hollow conductor, the part I_a returns inside and the rest outside. Since the total current enclosed by the inner surface of the given conductor is $-I_a$, and that enclosed by the outer surface is I_b, the magnetomotive intensity takes the values $-(I_a/2\pi a)$ and $(I_b/2\pi b)$, respectively, at these surfaces. This information is sufficient to determine the values of the constants A and B in the equation (59) governing current distribution. In fact, we

have

$$\begin{aligned} AI_1(\sigma a) + BK_1(\sigma a) &= -\frac{I_a}{2\pi a}, \\ AI_1(\sigma b) + BK_1(\sigma b) &= \frac{I_b}{2\pi b}, \end{aligned} \tag{71}$$

and therefore

$$\begin{aligned} A &= \frac{K_1(\sigma b)}{2\pi a D} I_a + \frac{I_1(\sigma a)}{2\pi b D} I_b, \\ B &= -\frac{I_1(\sigma b)}{2\pi a D} I_a - \frac{I_1(\sigma a)}{2\pi b D} I_b, \end{aligned} \tag{72}$$

where

$$D = I_1(\sigma b)K_1(\sigma a) - I_1(\sigma a)K_1(\sigma b). \tag{73}$$

Substituting these into the second equation of the set (59), we obtain the longitudinal electromotive intensity at any point of the conductor. We are interested, however, in its values at the surfaces since these values determine the surface impedances. Equating ρ successively to a and b, we obtain

$$\begin{aligned} E_z(a) &= Z_{aa}I_a + Z_{ab}I_b, \\ E_z(b) &= Z_{ba}I_a + Z_{bb}I_b, \end{aligned} \tag{74}$$

where [20]

$$\begin{aligned} Z_{aa} &= \frac{\eta}{2\pi a D}[I_0(\sigma a)K_1(\sigma b) + K_0(\sigma a)I_1(\sigma b)], \\ Z_{bb} &= \frac{\eta}{2\pi b D}[I_0(\sigma b)K_1(\sigma a) + K_0(\sigma b)I_1(\sigma a)], \\ Z_{ab} &= Z_{ba} = \frac{1}{2\pi g a b D}. \end{aligned} \tag{75}$$

The results embodied in equation (74) can be stated in the following two theorems:

Theorem 1: If the return path is wholly external ($I_a = 0$) or wholly internal ($I_b = 0$), the longitudinal electromotive intensity on that surface of a hollow conductor which is nearest to the return path equals the corresponding surface impedance per unit length multiplied by the total current flowing in the conductor; and the intensity on the other surface equals the transfer impedance per unit length multiplied by the total current.

[20] To obtain the last equation, it is necessary to use the identity

$$I_0(x)K_1(x) + K_0(x)I_1(x) = \frac{1}{x}.$$

Theorem 2: If the return path is partly external and partly internal the separate components of the intensity due to the two parts of the total current are calculated by the above theorem and then added to obtain the total intensities.

At high frequencies, or when the conductors are very large, (75) can be replaced by much simpler approximate expressions.[21] If, however, we are compelled to use the rigorous equations in numerical computations, it is convenient to express the Bessel functions in terms of Thomson functions. Two of these, the ber and bei functions, or Thomson functions of the first kind, have already been introduced. The functions of the second kind are defined in an entirely analogous fashion as

$$K_0(x\sqrt{i}) = \ker x + i \operatorname{kei} x. \tag{76}$$

Differentiating, we have

$$\sqrt{i}\, K_0'(x\sqrt{i}) = -\sqrt{i}\, K_1(x\sqrt{i}) = \ker' x + i \operatorname{kei}' x, \tag{77}$$

so that

$$K_1(x\sqrt{i}) = -\frac{\ker' x + i \operatorname{kei}' x}{\sqrt{i}}. \tag{78}$$

All these subsidiary functions have been tabulated;[22] but the process of computing the impedances is laborious nevertheless.

The Complex Poynting Vector[23]

In the preceding sections we have been able to determine the surface impedances of the coaxial conductors by reducing the field equations to the form of transmission line equations, and interpreting various terms accordingly. However, if the conductors are eccentric or of irregular shape, the *effective* surface impedances are more conveniently calculated by the use of the modified Poynting theorem.

This theorem states that, if E and H are the complex electromotive and magnetomotive intensities at any point, and if E^* and H^* are the conjugate complex numbers, then[24]

$$\int\int [EH^*]dS = g\int\int\int (EE^*)dv + i\omega\mu \int\int\int (HH^*)dv. \tag{79}$$

[21] See portion of this text under the heading "Approximate Formulæ for the Surface Impedance of Tubular Conductors," page 557.

[22] British Association Tables, 1912, pp. 57–68; 1915, pp. 36–38; 1916, pp. 108–122.

[23] For an early application of the Complex Poynting vector see Abraham v. Föppl, Vol. 1 (Ch. 3, Sec. 3).

[24] The brackets signify the vector product and the parentheses the scalar product of the vectors so enclosed. The inward direction of the normal to the surface is chosen as the positive direction. The division by 4π does not occur if the consistent practical system of units is used as it is done in this paper.

To get an insight into the significance of this equation, let us consider a conductor which is part of a single-mesh circuit, and extend our integrals over the region occupied by this conductor. Then the first integral on the right of (79) represents twice the power dissipated in heat in the conductor, while $\mu \int\int\int (HH^*)dv$ is four times the average amount of magnetic energy stored in it.

On the other hand, when we look at the conductor from the standpoint of circuit theory, these two quantities are respectively RI^2 and LI^2; R and L being by definition the "resistance" and "inductance" of the conductor. Hence we have the equation,

$$\int\int [EH^*]_n dS = (R + i\omega L)I^2 = ZI^2, \tag{80}$$

from which the *impedance* Z can be computed when the field intensities are known at the surface of the conductor.

If, on the other hand, the conductor is part of a two-mesh circuit and I_1 and I_2 are the amplitudes of the currents in meshes 1 and 2 respectively, the average amount of energy dissipated in heat per second can be regarded as made up of three parts, two of which are proportional to the squares of these amplitudes, while the third is proportional to their product. The first two of these parts being dependent on the magnitude of the current flowing in one mesh only are attributed to the *self-resistance* of the conductor to the corresponding current; the third part is attributed to the *mutual resistance* of the conductor. Designating the self-resistances by R_{11} and R_{22} and the mutual resistance by R_{12}, we represent the energy dissipated in heat in the form $1/2(R_{11}I_1^2 + 2R_{12}I_1I_2 + R_{22}I_2^2)$. Similarly, the average amount of energy stored in the conductor can be represented in the form $1/4(L_{11}I_1^2 + 2L_{12}I_1I_2 + L_{22}I_2^2)$, where L_{11} and L_{22} are called respectively *self-inductances* and L_{12} *mutual inductance*. In this case, equation (79) can be written as follows:

$$\int\int [EH^*]_n dS = Z_{11}I_1^2 + 2Z_{12}I_1I_2 + Z_{22}I_2^2, \tag{81}$$

where the quantities Z_{11}, Z_{22} and Z_{12} are respectively the *self-impedances* and the *mutual impedance* of the conductor.

In general, if the conductor is part of a k-mesh circuit, we can obtain all its self-and mutual impedances by evaluating the integral $\int\int [EH^*]_n dS$ over its surface, and picking out the coefficients of various combinations of I's.

We shall have an occasion to apply these results in computing the effect of eccentricity upon the resistance of parallel cylindrical conductors.

Approximate Formulæ for the Surface Impedance of Tubular Conductors

The exact formulæ (75) for the internal impedances of a tubular conductor are hard to use for numerical computations, but simple approximations can be easily obtained if the modified Bessel functions are replaced by their asymptotic expansions and the necessary division performed as far as the second term. Thus, we have

$$\begin{aligned} Z_{bb} &= \frac{\eta}{2\pi b}\left[\coth \sigma t + \frac{\pi}{2\sigma}\left(\frac{3}{a}+\frac{1}{b}\right)\right], \\ Z_{aa} &= \frac{\eta}{2\pi a}\left[\coth \sigma t - \frac{\pi}{2\sigma}\left(\frac{3}{b}+\frac{1}{a}\right)\right], \\ Z_{ab} &= \frac{\eta}{2\pi\sqrt{ab}}\operatorname{csch} \sigma t, \end{aligned} \tag{82}$$

where t is the thickness of the tube. Separating the real and imaginary parts, we have

$$\begin{aligned} R_{bb} &= \frac{1}{2b}\sqrt{\frac{\mu f}{\pi g}}\frac{\sinh u + \sin u}{\cosh u - \cos u} + \frac{a+3b}{16\pi g a b^2}, \\ R_{aa} &= \frac{1}{2a}\sqrt{\frac{\mu f}{\pi g}}\frac{\sinh u + \sin u}{\cosh u - \cos u} - \frac{b+3a}{16\pi g b a^2}, \\ R_{ab} &= \frac{1}{\sqrt{ab}}\sqrt{\frac{\mu f}{\pi g}}\frac{\sinh \frac{u}{2}\cos\frac{u}{2} + \cosh\frac{u}{2}\sin\frac{u}{2}}{\cosh u - \cos u}, \\ \omega L_{bb} &= \frac{1}{2b}\sqrt{\frac{\mu f}{\pi g}}\frac{\sinh u - \sin u}{\cosh u - \cos u}, \\ \omega L_{aa} &= \frac{1}{2a}\sqrt{\frac{\mu f}{\pi g}}\frac{\sinh u - \sin u}{\cosh u - \cos u}, \\ \omega L_{ab} &= \frac{1}{\sqrt{ab}}\sqrt{\frac{\mu f}{\pi g}}\frac{\sinh \frac{u}{2}\cos\frac{u}{2} - \cosh\frac{u}{2}\sin\frac{u}{2}}{\cosh u - \cos u} \\ |Z_{ab}| &= \frac{\sqrt{\mu f}}{\sqrt{\pi g a b(\cosh u - \cos u)}}, \end{aligned} \tag{83}$$

where $u = t\sqrt{2g\omega\mu}$.

It is obvious that in the equations for the self-resistances, the second terms represent the first corrections for curvature and vanish altogether if the conductors are plane. Although these formulæ were derived by using asymptotic expansions which are valid only when the argument is large, i.e., at high frequencies, the results are good even at low

frequencies, provided the tubular conductor is not too thick. Thus, if the frequency is 0, the first term in the above expression for R_{bb} becomes $1/2\pi gbt$ which is the d.-c. resistance of the tube if its curvature is neglected. The second term only partially corrects for curvature, the error being of the order of $t^2/8b^2$. Hence, if the thickness of the tube is not more than 25 per cent of its *high-frequency* radius, that is, the radius of the surface nearest the return path, the error is less than 1 per cent. The formula for the mutual impedance is exceedingly good down to zero frequency for all ordinary thicknesses.

If the frequency is very high, further approximations can be made and the formulæ simplified as follows:

$$
\begin{aligned}
R_{bb} &= \frac{1}{2b}\sqrt{\frac{\mu f}{\pi g}} + \frac{a+3b}{16\pi g a b^2}, \\
R_{aa} &= \frac{1}{2a}\sqrt{\frac{\mu f}{\pi g}} - \frac{b+3a}{16\pi g b a^2}, \\
R_{ab} &= \frac{\sqrt{2}}{\sqrt{ab}}\sqrt{\frac{\mu f}{\pi g}}\, e^{-(u/2)} \cos\left(u - \frac{\pi}{4}\right), \\
\omega L_{bb} &= \frac{1}{2b}\sqrt{\frac{\mu f}{\pi g}}, \\
\omega L_{aa} &= \frac{1}{2a}\sqrt{\frac{\mu f}{\pi g}}, \\
\omega L_{ab} &= \frac{\sqrt{2}}{\sqrt{ab}}\sqrt{\frac{\mu f}{\pi g}}\, e^{-(u/2)} \cos\left(u + \frac{\pi}{4}\right), \\
|Z_{ab}| &= \frac{\sqrt{2}}{\sqrt{ab}}\sqrt{\frac{\mu f}{\pi g}}\, e^{-(u/2)}.
\end{aligned}
\tag{84}
$$

If the ratio of the diameters of the tube is not greater than 4/3, then we have the following formula for the surface transfer impedance:

$$
\frac{|Z_{ab}|}{R_{\text{d.-c.}}} = \frac{u}{\sqrt{\cosh u - \cos u}}, \tag{85}
$$

which is correct to within 1 per cent at any frequency. This ratio is illustrated in Fig. 3. The ratios of the mutual resistance and the mutual reactance to the d.-c. resistance are shown in Fig. 4.

In the case of self-resistances, we let

$$
R_0 = \frac{1}{2\pi g r t}, \tag{86}
$$

where r is the high frequency radius of the tube. Thus R_0 is the d.-c. resistance of the tube if the curvature is neglected. Then we have approximately

$$\frac{R}{R_0} = \frac{u}{2} \cdot \frac{\sinh u + \sin u}{\cosh u - \cos u} \pm \frac{t}{2r}, \tag{87}$$

if the tube is fairly thin. The curvature correction is positive if the

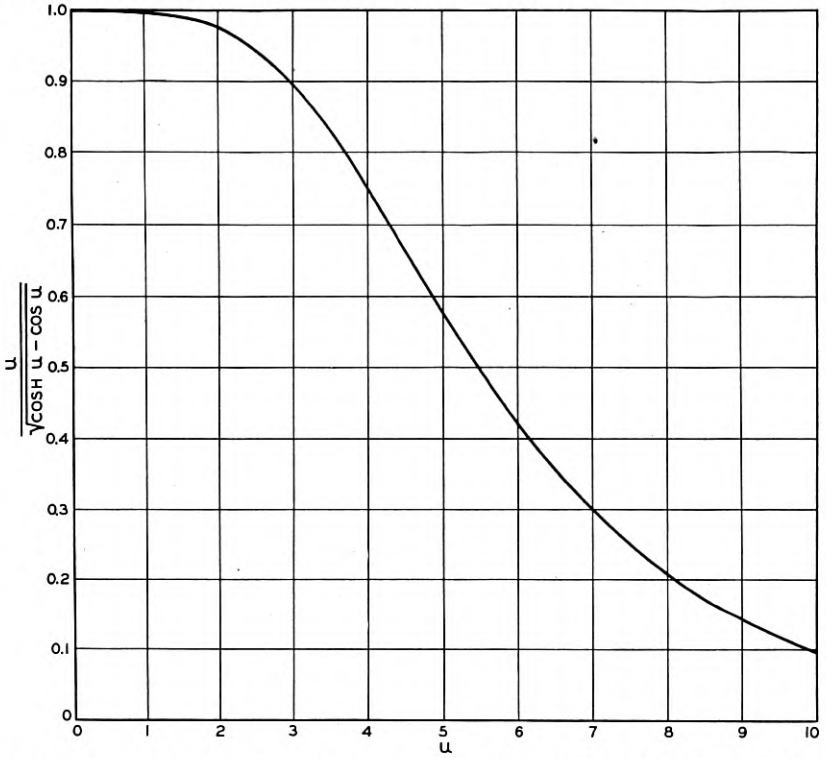


Fig. 3—The transfer impedance from one surface of a cylindrical shell to the other. The curve represents its ratio to the d-c. resistance.

return path is external, and negative if it is internal. The graph of the first term is shown in Fig. 5.

An interesting observation can be made at once from the formulæ (83) for the self-resistances of a tubular conductor. If the frequency is kept fixed and the thickness of the conductor is increased from 0, its resistance (with either return) passes through a sequence of maxima and minima.[25] The first minimum occurs when $u = \pi$, i.e., when

[25] The general fluctuating character of this function was noted by Mrs. S. P. Mead [12].

$t = \sqrt{\pi}/(2\sqrt{g\mu f})$; the first maximum occurs when $u = 2\pi$, etc. This fluctuation in resistance is due to the phase shift in the current density as we proceed from the surface of the conductor to deeper layers. The "optimum" resistance is $R_0((\pi/2) \tanh \pi/2) = 1.44R_0$, plus or

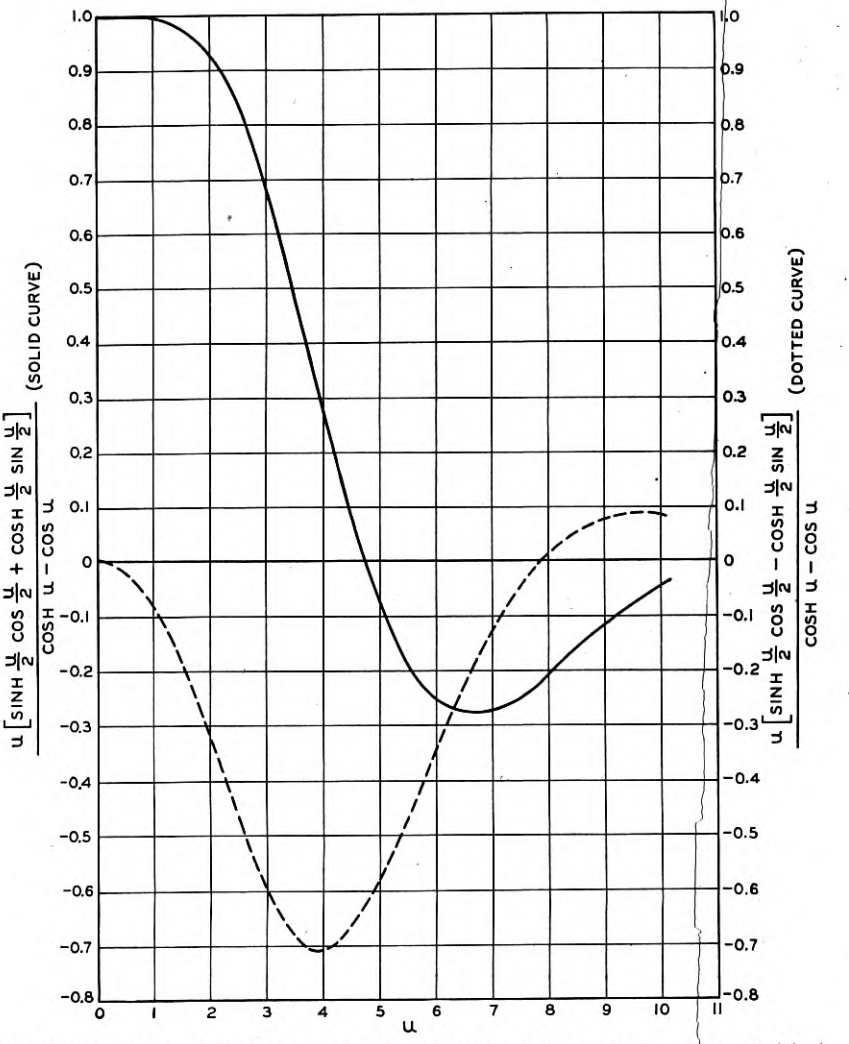


Fig. 4—The ratios of the transfer resistance and transfer reactance of a cylindrical shell to its d-c. resistance.

minus the curvature correction $t/2r$. If curvature is disregarded, the ratio of the optimum resistance to the resistance of the infinitely thick conductor with the same internal diameter as the hollow con-

ductor is $\tanh \pi/2 = 0.92$. When $u = 2\pi$, the ratio reaches its first maximum $\coth \pi = 1.004$. At 1 megacycle the optimum thickness of a copper conductor is about 0.1038 mm.

By a method of successive approximations, H. B. Dwight has ob-

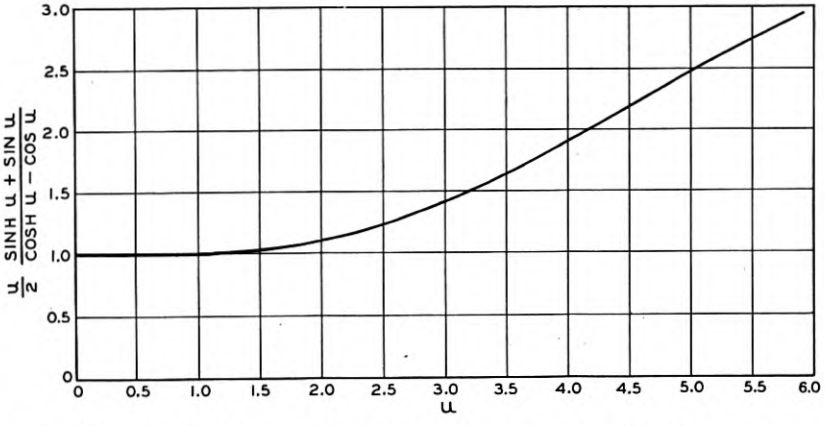


Fig. 5—The skin effect in cylindrical shells. The curve represents the ratio of the a-c. resistance of a typical shell to its d-c. resistance.

tained the impedance of a tubular conductor with an external coaxial return.[26] His final results appear as the ratio of two infinite power series, which converge for all values of the variables involved, though they can be used advantageously in numerical computations only when the frequencies are fairly low and the convergence is rapid. We shall merely indicate how Dwight's formula and other similar formulæ can be obtained directly from the exact equations (75).

Let us replace the outer radius b of (75) by $a + t$, where t is the thickness of the wall, and replace the various Bessel functions by their Taylor series in t:

$$\begin{aligned} I_0(\sigma b) &= I_0(\sigma a + \sigma t) = \sum_{n=0}^{\infty} \frac{(\sigma t)^n}{n!} I_0^{(n)}(\sigma a), \\ K_0(\sigma b) &= K_0(\sigma a + \sigma t) = \sum_{n=0}^{\infty} \frac{(\sigma t)^n}{n!} K_0^{(n)}(\sigma a), \\ I_1(\sigma b) &= I_0'(\sigma b) = \sum_{n=0}^{\infty} \frac{(\sigma t)^n}{n!} I_0^{(n+1)}(\sigma a), \\ -K_1(\sigma b) &= K_0'(\sigma b) = \sum_{n=0}^{\infty} \frac{(\sigma t)^n}{n!} K_0^{(n+1)}(\sigma a). \end{aligned} \tag{88}$$

[26] "Skin Effect in Tubular and Flat Conductors," *A. I. E. E. Journal*, Vol. 37 (1918), p. 1379.

We thus obtain

$$Z_{bb} = \frac{2\eta}{b} \frac{\sum_{n=0}^{\infty} A_n \frac{(\sigma t)^n}{n!}}{\sum_{n=0}^{\infty} A_{n+1} \frac{(\sigma t)^n}{n!}}, \tag{89}$$

where A_n is defined as

$$A_n = \begin{vmatrix} I_0'(\sigma a) & I_0^{(n)}(\sigma a) \\ K_0'(\sigma a) & K_0^{(n)}(\sigma a) \end{vmatrix}. \tag{90}$$

In spite of the complicated appearance of (90) the A's are in reality very simple functions of σa, as the accompanying list (91) will show.[27]

$$\begin{aligned} A_0 &= \frac{1}{\sigma a}, \quad A_1 = 0, \quad A_2 = \frac{1}{\sigma a}, \quad A_3 = -\frac{1}{\sigma^2 a^2}, \\ A_4 &= \frac{1}{\sigma a} + \frac{3}{\sigma^3 a^3}, \quad A_5 = -\frac{2}{\sigma^2 a^2} - \frac{12}{\sigma^4 a^4}, \\ A_6 &= \frac{1}{\sigma a} + \frac{9}{\sigma^3 a^3} + \frac{60}{\sigma^5 a^5}. \end{aligned} \tag{91}$$

The formula (90) can be made more rapidly convergent by partially summing the numerator and the denominator by means of hyperbolic functions. Thus, the numerator becomes

$$\sqrt{\frac{a}{b}} \cosh \sigma t + \frac{\sinh \sigma t}{2\sigma a}\left[1 - \frac{3t}{4a} + \cdots\right] - \frac{3t^2}{a^2} + \cdots,$$

and the denominator

$$\sqrt{\frac{a}{b}} \sinh \sigma t + \frac{3(\sigma t \cosh \sigma t - \sinh \sigma t)}{8(\sigma a)^2} + \cdots.$$

The reader will readily see that there would be no difficulty in using this method to obtain other expansions somewhat similar to (89). For example, we might write $a = b - t$ in (75) and express our results in terms of the outer radius. In this respect the method that we have used has greater flexibility than Dwight's; but there seems to be little advantage gained from it, since the simple formulæ (82) are sufficient for most practical purposes.

[27] The values given in (91) are *exact*, not approximate. One of them, namely,

$$A_1 = \begin{vmatrix} I_0'(\sigma a) & I_0(\sigma a) \\ K_0'(\sigma a) & K_0(\sigma a) \end{vmatrix} \equiv \frac{1}{\sigma a},$$

is one of the fundamental identities found in all books on Bessel functions. The rest are consequences of analogous, though ess familiar, identities. The general expressions for the coefficients A_n were obtained by H. Pleijel [20].

Internal Impedances of Laminated Conductors

So far we have supposed that all conductors were homogeneous. We shall now consider a somewhat more general conductor composed of n coaxial layers of different substances. As before, we are interested in finding expressions for the internal impedances; besides, we may wish to know how the total current is distributed between the different layers of the conductor.

To begin with, let us suppose that a coaxial return path is provided *outside* the given conductor. We number our layers consecutively and call the inner layer the first. Let $Z_{aa}^{(m)}$ and $Z_{bb}^{(m)}$ be the surface impedances of the mth layer, the first when the return is internal, the other when it is external; and let $Z_{ab}^{(m)}$ be the transfer impedance from one surface to the other. Formulæ for these impedances have already been obtained in the section under "The Surface Impedances of Hollow Cylindrical Shells," page 552. Also, let $z_{bb}^{(m)}$ be the *surface impedance of the first m layers* with external return; that is, the ratio of the longitudinal electromotive intensity at the outer surface of the mth layer to the total current I_m in all m layers.[28]

By hypothesis, there is no return path inside the laminated conductor as a whole. Hence, when we fix our attention on any one layer alone, say the mth, we may say that the current *in this layer* returns partly through the $m - 1$ layers within it, and partly outside. In the $m - 1$ inner layers, however, the current is assumed to be I_{m-1} in the outward direction—or what amounts to the same thing $-I_{m-1}$ in the return direction. Hence we conclude that, of the current $I_m - I_{m-1}$ in the layer under discussion, I_m returns *outside* and $-I_{m-1}$ *inside*. Substituting these values in Theorem 2 on page 555, we find that the electromotive intensity along the inner surface of the layer is $Z_{ab}^{(m)} I_m - Z_{aa}^{(m)} I_{m-1}$.

But the *inner* surface of the mth layer is the *outer* surface of the composite conductor comprising the $m - 1$ inner layers, and by Theorem 1, the electromotive intensity on this outer surface is $z_{bb}^{(m-1)} I_{m-1}$. As the two must be equal, we obtain an equation from which we can determine the ratio of the current flowing in the first $m - 1$ layers to that flowing in m layers. This is

$$\frac{I_{m-1}}{I_m} = \frac{Z_{ab}^{(m)}}{Z_{aa}^{(m)} + z_{bb}^{(m-1)}}. \tag{92}$$

In this formula for the effect of an extra layer on the current dis-

[28] In this notation, the current flowing in the mth layer is $I_m - I_{m-1}$. It should also be noted that $Z_{bb}^{(1)} = z_{bb}^{(1)}$.

tribution, it will be noted that the denominator is the impedance (with internal return) of the added layer plus the original impedance.

We now consider the electromotive intensity on the outer surface of the mth layer, which is $z_{bb}^{(m)} I_m$ on the one hand, and $(Z_{bb}^{(m)} I_m - Z_{ab}^{(m)} I_{m-1})$ on the other. Thus, we have the following equation,

$$z_{bb}^{(m)} = Z_{bb}^{(m)} - Z_{ab}^{(m)} \frac{I_{m-1}}{I_m} = Z_{bb}^{(m)} - \frac{[Z_{ab}^{(m)}]^2}{Z_{aa}^{(m)} + z_{bb}^{(m-1)}}, \quad (93)$$

expressing the effect of an additional layer upon the impedance of the conductor.

This equation is a convenient reduction formula. Starting with the first layer (for which $z_{bb}^{(1)} = Z_{bb}^{(1)}$), we add the remaining layers one by one and thus obtain the impedance of the complete conductor in the form of the following continued fraction:

$$z_{bb}^{(n)} = Z_{bb}^{(n)} - \frac{[Z_{ab}^{(n)}]^2}{Z_{aa}^{(n)} + Z_{bb}^{(n-1)} + \dfrac{[Z_{ab}^{(n-1)}]^2}{Z_{aa}^{(n-1)} + Z_{bb}^{(n-2)} + \cdots \dfrac{[Z_{ab}^{(2)}]^2}{Z_{aa}^{(2)} + Z_{bb}^{(1)}}}} \quad (94)$$

We can also get a reduction formula for the transfer impedance between the inner and outer surfaces of the composite conductor formed by the first m layers. To do so, it is only necessary to note that, since the inner surface of the first $m - 1$ layers is also the inner surface of the first m layers as well, the electromotive intensity on that surface can be expressed either as $z_{ab}^{(m-1)} I_{m-1}$ or as $z_{ab}^{(m)} I_m$. Thus, we have

$$z_{ab}^{(m)} = z_{ab}^{(m-1)} \frac{I_{m-1}}{I_m} = \frac{Z_{ab}^{(m)} z_{ab}^{(m-1)}}{Z_{aa}^{(m)} + z_{bb}^{(m-1)}}. \quad (95)$$

By noting that $z_{ab}^{(1)} = Z_{ab}^{(1)}$, we can determine successively the transfer impedances across the first two layers, the first three, and so on. This formula is not quite as simple as (94), owing to the presence of $z_{bb}^{(m-1)}$ in its denominator, and it is therefore not expedient to evaluate $z_{ab}^{(m)}$ explicitly; but it is not prohibitively cumbersome from the numerical standpoint when the computations are made step by step.

Although in deducing equations (93) and (95) we supposed that the added layer was homogeneous, the equations are correct even if this layer consists of several coaxial layers, provided $Z_{aa}^{(m+1)}$ and $Z_{ab}^{(m+1)}$ are interpreted as the impedances of the added non-homogeneous layer in the absence of the original core of m layers. These latter

impedances themselves have to be computed by means of equations (94) and (95).

If the return path is inside the laminated conductor, instead of outside, formulæ (92) and (93) still hold, provided we interchange a and b, and count layers from the outside instead of the inside, so that $m = 1$ is the outermost, rather than the innermost, layer.

The basic rule for determining the surface impedances of laminated conductors can be put into the following verbal form:

Theorem 3: Let two conductors, both of which may be made up of coaxial layers, fit tightly one inside the other. Any surface self-impedance of the compound conductor equals the individual impedance of the conductor nearest to the return path diminished by the fraction whose numerator is the square of the transfer impedance across this conductor and whose denominator is the sum of the surface impedances of the two component conductors if each is regarded as the return path for the other. The transfer impedance of the compound conductor is the fraction whose numerator is the product of the transfer impedances of the individual conductors and whose denominator is that of the self-impedance.

If two coaxial conductors are short-circuited at intervals, short compared to the wave-length, the above theorem holds even if the conductors do not fit tightly one over the other, provided we add in the denominators a third term representing the inductive reactance of the space between the conductors.

Disks as Terminal Impedances for Coaxial Pairs

So far we have been concerned only with infinitely long pairs. We now take up a problem of a different sort; namely, the design of a disk which, when clapped on the end of such a pair, will not give rise to a reflected wave.

The line of argument will be as follows: To begin with, we shall assume a disk of arbitrary thickness h, compute the field which will be set up in it, and then adjust the thickness so as to make this field match that which would exist in the dielectric of an infinite line.

The field in the disk has to satisfy equation (2) where $i\omega\epsilon$ can be disregarded by comparison with g. Thus, we have

$$\frac{\partial H_\varphi}{\partial z} = -gE_\rho, \qquad \frac{1}{\rho}\frac{(\rho H_\varphi)}{\partial \rho} = gE_z,$$
$$\frac{\partial E_z}{\partial \rho} - \frac{\partial E_\rho}{\partial z} = i\omega\mu H_\varphi. \tag{96}$$

In the dielectric between the coaxial conductors, the longitudinal displacement current density is very small; in fact, it would be zero if the conductors were perfect. This current density is continuous across the surface of the disk and, therefore, gE_z is exceedingly small. Hence, the second of the above equations becomes approximately

$$\frac{\partial(\rho H_\varphi)}{\partial \rho} = 0; \tag{97}$$

so that

$$H_\varphi = \frac{P}{\rho}, \tag{98}$$

where P is independent of ρ but may be a function of z. Under these conditions, the remaining two equations are

$$\frac{\partial E_\rho}{\partial z} = -i\omega\mu H_\varphi, \qquad \frac{\partial H_\varphi}{\partial z} = -gE_\rho. \tag{99}$$

From the form of these equations and from (98), we conclude that the general expressions for the intensities in the disk are

$$H_\varphi = \frac{Ae^{\sigma z} + Be^{-\sigma z}}{\rho}, \qquad E_\rho = \frac{\sigma[Be^{-\sigma z} - Ae^{\sigma z}]}{g\rho}, \tag{100}$$

where $\sigma = \sqrt{g\omega\mu i}$.

On the outside flat surface of the disk (given by $z = h$ where h is the thickness of the plate), the magnetomotive intensity is very nearly zero;[29] therefore,

$$Ae^{\sigma h} + Be^{-\sigma h} = 0. \tag{101}$$

From this we obtain

$$A = -Ce^{-\sigma h}, \qquad B = Ce^{\sigma h}, \tag{102}$$

where C is some constant. Thus equations (100) can be written as follows:

$$H_\varphi = \frac{C \sinh \sigma(h - z)}{\rho},$$

$$E_\rho = \frac{\sigma C \cosh \sigma(h - z)}{g\rho}; \tag{103}$$

and at the boundary between the disk and the dielectric of the transmission line ($z = 0$), we have

$$\frac{E_\rho}{H_\varphi} = \frac{\sigma}{g} \coth \sigma h. \tag{104}$$

[29] On account of the negligibly small longitudinal current in the disk.

On the other hand, if there is to be no reflection this must equal $\sqrt{\mu/\epsilon}$ by equation (24). Hence

$$\frac{\sigma}{g}\coth \sigma h = \sqrt{\frac{\mu}{\epsilon}}. \tag{105}$$

If σh is small, coth σh equals approximately $1/\sigma h$, and

$$h = \frac{1}{g}\sqrt{\frac{\epsilon}{\mu}}\,\text{cm}. \tag{106}$$

Under these conditions, the generalized flux of energy across the inner surface of the disk is, in accordance with the text under "The Complex Poynting Vector," page 555, and equation (14),

$$\int_0^{2\pi}\int_{b'}^{a''} E_\rho H_\varphi^* \rho\, d\rho\, d\varphi = \frac{1}{2\pi}\sqrt{\frac{\mu}{\epsilon}}\log\frac{a''}{b'}\, I^2. \tag{107}$$

Thus, the impedance of this disk is a pure resistance equal to the characteristic impedance of the coaxial pair.

Cylindrical Waves and the Problem of Cylindrical Shields [30]

It is well known that when two transmission lines are side by side, to a greater or lesser extent they interfere with each other. This interference is usually analyzed into "electromagnetic crosstalk" and "electrostatic crosstalk."

Thus, electric currents in a pair of parallel wires produce a magnetic field with lines of force perpendicular to the wires. These lines cut the other pair of wires and induce in them electromotive forces, thereby producing what is usually called the "electromagnetic crosstalk"; this crosstalk is seen to be proportional to the current flowing in the first pair. The "electrostatic crosstalk," on the other hand, is caused by electric charges induced on the wires of the second system; these charges are proportional to the potential difference existing between the wires of the "disturbing" transmission line.

The distinction between two types of crosstalk is valid, although the terminology is somewhat unfortunate; the word "electromagnetic" is used in too narrow a sense and the word "electrostatic" is a contradiction in terms since electric currents and charges in a transmission line are variable. The terms "impedance crosstalk" and "admittance crosstalk" would be preferable because the former is due to a distributed mutual series impedance between two lines and the latter is produced by a distributed mutual shunt admittance.

[30] Since this paper was written, a related paper has been published by Louis V. King [18]. However, the physical picture here developed appears to be new. The earliest writer who treated the problem of electromagnetic shielding is H. Pleijel [21].

The crosstalk between two parallel pairs (this applies to twisted pairs as well) can be reduced by enclosing each pair in a cylindrical metallic shield. It is the object of this and the following two sections to develop a theory for the design of such shields.

This theory is based upon an assumption that in so far as the radial movement of energy toward and away from the wires is concerned we can disregard the non-uniform distribution of currents and charges along the length of the wires. No serious error is introduced thereby as long as the radius of the shield is small by comparison with the wave-length. The field around the wires is considered, therefore, as due to superposition of two two-dimensional fields of the types given by equations (4) and (5).

The actual computation of the effectiveness of a given shield will be reduced to an analogous problem in Transmission Line Theory.

Equations (4) and (5) are too general as they stand. Strictly speaking the effect of a shield upon an arbitrary two-dimensional field cannot be expressed by a single number. The field at various points outside the shield will be reduced by it in different ratios. However, any such field can be resolved into "cylindrical waves," each of which is reduced by the shield everywhere in the same ratio. Moreover, to all practical purposes the field produced by electric currents (or electric charges) in a pair of wires is just such a pure cylindrical wave.

Since both E and H are periodic functions of the coordinate φ, they can be resolved into Fourier series. The name "cylindrical waves" will be applied to the fields represented by the separate terms of the series. As the name indicates the wave fronts of these waves are cylindrical surfaces, although owing to relatively low frequencies and long wave-lengths used in practice the progressive motion of these waves is not clearly manifested except at great distances from the wires.

Turning our attention specifically to magnetic cylindrical waves *of the nth order*, and writing the field components tangential to the wave fronts in the form $E \cos n\varphi$ and $H \cos n\varphi$, we have from equations (5):

$$\frac{dE}{d\rho} = -i\omega\mu H, \quad \frac{d(\rho H)}{d\rho} = -\left[(g + i\omega\epsilon)\rho + \frac{n^2}{i\omega\mu\rho}\right]E. \quad (108)$$

From these we obtain

$$\rho^2\frac{d^2E}{d\rho^2} + \rho\frac{dE}{d\rho} = [i\omega\mu(g + i\omega\epsilon)\rho^2 + n^2]E. \quad (109)$$

This equation, being of the second order, possesses two independent solutions: one for diverging cylindrical waves and the other for reflected waves. The ratio of E to H in the first case and its negative in the second will be called the *radial impedance* offered by the medium to cylindrical waves.

In the next section we shall determine radial impedances in dielectrics and metals and show that for all practical purposes the attenuation of cylindrical waves in metals is exponential. The significance of the radial impedance is the same as that of the characteristic impedance of a transmission line. When a cylindrical wave passes from one medium into another, a *reflection* takes place unless the radial impedances are the same in the two media. Thus if E_0 and H_0 are the *impressed* intensities (at the boundary between the two media), E_r and H_r the reflected and E_t and H_t, the transmitted intensities, we have

$$E_0 + E_r = E_t \qquad \text{and} \qquad H_0 + H_r = H_t, \tag{110}$$

since both intensities must be continuous. On the other hand, if k is the ratio of the impedance in the first medium to that in the second, then equations (110) become

$$kH_0 - kH_r = H_t \qquad \text{and} \qquad H_0 + H_r = H_t. \tag{111}$$

Solving we obtain

$$H_t = \frac{2k}{k+1} H_0 \qquad \text{and} \qquad E_t = \frac{2}{k+1} E_0. \tag{112}$$

The reflection loss will be defined as

$$R = 20 \log_{10} \frac{|H_0|}{|H_t|} \text{ decibels.} \tag{113}$$

When a wave passes through a shield, it encounters two boundaries and if the shield is electrically thick, that is, if the attenuation of the wave in the shield is so great that secondary reflections can be disregarded without introducing a serious error, the total reflection loss is the sum of the losses at each boundary. The first loss can be computed directly from (112) and the second from the same equation if we replace k by its reciprocal. Thus, the total reflection loss for electrically thick shields is

$$R = 20 \log_{10} \frac{|k+1|^2}{4|k|} \text{ decibels.} \tag{114}$$

When the ratio of the impedances is very large by comparison with unity, the formula becomes

$$R = 20 \log_{10} \frac{|k|}{4}, \tag{115}$$

and when k is very small, then

$$R = 20 \log_{10} \frac{1}{4|k|}. \tag{116}$$

In the next section we shall see that to all practical purposes, the wave in the shield is attenuated exponentially. If α is the attenuation constant in nepers and if t is the thickness of the shield, then the attenuation loss is

$$A = 8.686 \alpha t \text{ decibels} \tag{117}$$

and the total reduction in the magnetomotive intensity due to the presence of the shield is

$$S = R + A. \tag{118}$$

The electromotive intensity is reduced in the same ratio.

But if the shield is not electrically thick, a correction term has to be added to the reflection loss. This correction term can be shown to be [31]

$$C = 20 \log_{10} \left| 1 - \frac{(k-1)^2}{(k+1)^2} e^{-2\Gamma t} \right| \text{ decibels}, \tag{119}$$

and if k is very large or very small by comparison with unity then

$$C \doteq 3 - 8.686 \alpha t + 10 \log_{10} (\cosh 2\alpha t - \cos 2\beta t). \tag{120}$$

Equation (120) does not hold down to $t = 0$; when Γt is nearly zero, then

$$C = 20 \log_{10} \left| 1 - \frac{(k-1)^2}{(k+1)^2} \right|. \tag{121}$$

So far we supposed that the shields were coaxial with the source. If this is not so, it is always possible to replace any given line source within the shield by an equivalent system of line sources coaxial with the shield and emitting cylindrical waves of proper orders. Mathematically this amounts to a change of the origin of the coordinate system. In the next section we shall see that the shielding effectiveness is not the same for all cylindrical waves. This means, of course, that if the shield is not coaxial with the source, the total reduction in

[31] Here, $\Gamma = \alpha + i\beta$ is the propagation constant in the shield.

the field depends upon the position of the measuring apparatus. The variation is very small, however, unless the source is almost touching the shield and it can be stated that approximately the shielding effectiveness is independent of the position of the source.

It is interesting to observe from the accompanying tables that while the attenuation loss is greater in iron than in copper, the reflection loss is greater at a copper surface. In fact, at some frequencies the impedances of iron and air nearly match and practically no reflection takes place. Hence, a thin copper shield may be more effective than an equally thin iron shield. And if a composite shield is made of copper and iron, the shield will be more effective if copper layers are placed on the outside to take advantage of the added reflection.

TABLE I

THE ABSOLUTE VALUE OF THE RADIAL IMPEDANCE OFFERED BY AIR TO CYLINDRICAL MAGNETIC WAVES OF THE FIRST ORDER (IN MICROHMS)

f	Radius = 0.5 cm.	1 cm.	2 cm.
1 cycle	0.0395	0.07896	0.1579
10 cycles	0.395	0.790	1.58
100 cycles	3.95	7.90	15 8
1 kilocycle	39.5	79.0	158.
10 kilocycles	395.	790.	1,580.
100 kilocycles	3,950.	7,900.	15,800.
1 megacycle	39,500.	79,000.	158,000.
10 megacycles	395,500.	790,000.	1.58 ohms
100 megacycles	3.95 ohms	7.9 ohms	15.8 ohms

TABLE II

THE INTRINSIC IMPEDANCE OF CERTAIN METALS $(\eta)/(\sqrt{i})$ IN MICROHMS

f	Copper $g = 5.8005 \times 10^5$ mhos/cm. $\mu = 0.01257$ μh/cm.	Lead $g = 4.8077 \times 10^4$ mhos/cm. $\mu = 0.01257$ μh/cm.	Aluminum $g = \frac{1}{3} 10^6$ mhos/cm. $\mu = 0.01257$ μh/cm.	Iron $g = 10^5$ mhos/cm. $\mu = 1.257$ μh/cm. = (100 relative to copper)
1 cycle	0.369	1.28	0.487	8.88
10 cycles	1.17	4.05	1.54	28.1
100 cycles	3.69	12.8	4.87	88.8
1 kilocycle	11.7	40.5	15.4	281.
10 kilocycles	36.9	128.	48.7	888.
100 kilocycles	117.	405.	154.	2,810.
1 megacycle	369.	1,280.	487.	8,880.
10 megacycles	1,170.	4,050.	1,540.	28,100.
100 megacycles	3,690.	12,800.	4,870.	88,800.

TABLE III

THE INTRINSIC PROPAGATION CONSTANT OF CERTAIN METALS

f	Copper $g = 5.8005 \times 10^5$ mhos cm. $\mu = 0.01257$ μh/cm.		Lead $g = 4.8077 \times 10^4$ mhos/cm. $\mu = 0.01257$ μh/cm.		Aluminum $g = \frac{1}{3}10^6$ mhos/cm. $\mu = 0.01257$ μh/cm.		Iron $g = 10^5$ mhos/cm. $\mu = 1.257$ μh/cm. = (100 relative to copper)	
	$\frac{\sigma}{\sqrt{i}}$ in $\frac{\text{nepers}}{\text{cm.}}$	$\frac{\sigma}{\sqrt{i}}$ in db/cm.	$\frac{\sigma}{\sqrt{i}}$ in $\frac{\text{nepers}}{\text{cm.}}$	$\frac{\sigma}{\sqrt{i}}$ in db/cm.	$\frac{\sigma}{\sqrt{i}}$ in $\frac{\text{nepers}}{\text{cm.}}$	$\frac{\sigma}{\sqrt{i}}$ in db/cm.	$\frac{\sigma}{\sqrt{i}}$ in $\frac{\text{nepers}}{\text{cm.}}$	$\frac{\sigma}{\sqrt{i}}$ in db/cm.
1 cycle	0.214	1.86	0.0616	.535	0.162	1.41	0.888	7.72
10 cycles	0.677	5.88	0.195	1.69	0.513	4.46	2.81	24.4
100 cycles	2.14	18.6	0.616	5.35	1.62	14.1	8.88	77.2
1 kilocycle	6.77	58.8	1.95	16.9	5.13	44.6	28.1	244.
10 kilocycles	21.4	186.	6.16	53.5	16.2	141.	88.8	772.
100 kilocycles	67.7	588.	19.5	169.	51.3	446.	281.	2,440.
1 megacycle	214.	1,860.	61.6	535.	162.	1,410.	888.	7,720.
10 megacycles	677.	5,880.	195.	1690.	513.	4,460.	2810.	24,400.
100 megacycles	2140.	18,600.	616.	5350.	1620.	14,100.	8880.	77,200.

Cylindrical Waves in Dielectrics and Metals

In good dielectrics g is small by comparison with $\omega\epsilon$ and the first term on the right in (109) very nearly equals $(2\pi\rho/\lambda)^2$ where λ is the wave-length. But we are interested in wave-lengths measured in miles and shields with diameters measured in inches; thus we shall write (109) in the following approximate form:

$$\rho^2 \frac{d^2E}{d\rho^2} + \rho \frac{dE}{d\rho} - n^2E = 0. \tag{122}$$

When $n \neq 0$ there are two independent solutions

$$E_1 = \rho^{-n} \quad \text{and} \quad E_2 = \rho^n; \tag{123}$$

and when $n = 0$,

$$E_1 = \log \rho \quad \text{and} \quad E_2 = 1. \tag{124}$$

The corresponding expressions for H are, by (108),

$$H_1 = \frac{n\rho^{-n-1}}{i\omega\mu} \quad \text{and} \quad H_2 = -\frac{n\rho^{n-1}}{i\omega\mu}, \tag{125}$$

in the first case, and

$$H_1 = \frac{1}{i\omega\mu\rho} \quad \text{and} \quad H_2 = 0, \tag{126}$$

in the second.

The second case in which E_1 and H_1 are the electromotive and magnetomotive intensities in the neighborhood of an isolated wire carrying electric current is of interest to us only in so far as it helps to interpret (123) and (125). If we were to consider $2n$ infinitesimally thin wires equidistributed upon the surface of an infinitely narrow cylinder, the adjacent wires carrying equal but oppositely directed currents of strength sufficient to make the field different from zero, and calculate the field, we should obtain expressions proportional to E_1 and H_1. An actual cluster of $2n$ wires close together would generate principally a cylindrical wave of order n; the strengths of other component waves of order $3n$, $5n$, etc. rapidly diminish as the distance from the cluster becomes large by comparison with the distance between the adjacent wires of the cluster. For the purposes of shielding design we can regard a pair of wires as generating a cylindrical wave of the first order ($n = 1$). The radial impedance of an nth order wave is

$$Z_\rho = \frac{E_1}{H_1} = \frac{i\omega\mu\rho}{n}, \tag{127}$$

and that of the corresponding reflected wave has the same value. It should be noted that by the "reflected" cylindrical wave in the space enclosed by a shield, we mean the sum total of an infinite number of successive reflections. Each of the latter waves condenses on the axis and diverges again only to be re-reflected back; in a steady state all these reflected waves interfere with each other and form what might be called a "stationary reflected wave." Not being interested in any other kind of reflected waves we took the liberty of omitting the qualification.

In conductors the attenuation of a wave due to energy dissipation is much greater (except at extremely low frequencies) than that due to the cylindrical divergence of the wave. Hence, in the shield we can regard the wave as plane and write (108) in the following approximate form:

$$\frac{dE}{d\rho} = -i\omega\mu H, \qquad \frac{dH}{d\rho} = -gE. \tag{128}$$

In form, these are exactly like ordinary transmission line equations. Hence, in a shield the radial impedance is simply the intrinsic impedance of the metal,

$$Z_\rho = \eta = \sqrt{\frac{i\omega\mu}{g}} \text{ ohms}, \tag{129}$$

and the propagation constant,

$$\sigma = \sqrt{i\omega\mu g} = \sqrt{\pi f\mu g}\,(1 + i) \text{ nepers/cm}. \tag{130}$$

The exact value of the radial impedance in metals can be found by solving (108). Thus, we can obtain

$$Z_\rho = -\eta \frac{K_n(\sigma\rho)}{K_n'(\sigma\rho)} \tag{131}$$

for diverging waves, and

$$Z_\rho = \eta \frac{I_n(\sigma\rho)}{I_n'(\sigma\rho)} \tag{132}$$

for the reflected waves.

Cylindrical waves of the electric type can be treated in the same manner. It turns out that the transmission laws in metals are identical with those for magnetic waves. The radial impedance in perfect dielectrics, on the other hand is given by

$$Z_\rho = \frac{n}{i\omega\epsilon\rho}. \tag{133}$$

This is enormous by comparison with the impedance in metals, thereby explaining an almost perfect "electrostatic" shielding offered by metallic substances. Even when the frequency is as high as 100 kc. the radial impedance of air 1 cm. from the source is about 36×10^6 ohms while the impedance of a copper shield is only 117×10^{-6} ohms. The reflection loss is approximately 220 db.

Power Losses in Shields

As we have shown in the text under "The Complex Poynting Vector," page 555, the average power dissipated in a conductor is the real part of the integral $\Phi = 1/2 \int\int [EH^*]_n dS$ taken over the surface of the conductor. If the source of energy is inside a shield, the integration need be extended only over its inner surface, because the average energy flowing outward through this surface is almost entirely dissipated in the shield, the radiation loss being altogether negligible. If a cylindrical wave whose intensities at the inner surface of the shield of radius "a" are

$$H_\varphi = H_0 \cos n\varphi, \qquad H_\rho = H_0 \sin n\varphi, \qquad E_z = \eta H_\varphi, \tag{134}$$

$\eta = i\omega\mu a/n$ being the radial impedance in the dielectric, is impressed upon the inner surface of the shield, a reflected wave is set up. The resultant of the magnetomotive intensities in the two is readily found to be $(2k/k + 1)H_0$, where k is the ratio of the radial impedance of the dielectric column inside the shield to the impedance Z looking into the shield. If the shield is electrically thick, the impedance Z is obviously the radial impedance of the shield; otherwise it is modified somewhat by reflection from the outside of the shield. The average power loss in the shield per centimeter of length is, then, the real part of

$$\Phi = \frac{2\pi a k k^* Z}{(k + 1)(k^* + 1)} H_0 H_0^*. \tag{135}$$

This becomes simply

$$\Phi = 2\pi a Z H_0 H_0^*, \tag{136}$$

if the frequency is so high that k is large as compared with unity.

If the source of the impressed field is a pair of wires along the axis of the shield, the magnetomotive intensity on the surface of the shield can be shown to be

$$H_\varphi = \frac{l}{2\pi a^2} I \cos \varphi, \tag{137}$$

where l is the separation between the axes of the wires. Therefore,

$$\Phi = \frac{kk^*l^2Z}{2\pi a^3(k+1)(k^*+1)} I^2. \tag{138}$$

Resistance of Nearly Coaxial Tubular Conductors

When two tubular conductors are not quite coaxial, a proximity effect [32] appears which disturbs the symmetry of current distribution and therefore somewhat increases their resistance. This effect can be estimated by the following method of successive approximations. To begin with, we assume a symmetrical current distribution in the *inner* conductor. The magnetic field outside this conductor is then the same as that of a simple source along its axis and can be replaced by an equivalent distribution of sources situated along the axis of the outer conductor. The principal component of this distribution is a simple source of the same strength as the actual source and does not enter into the proximity effect. The next largest component is a double source given by

$$\begin{aligned} E_z &= \frac{i\omega\mu lI}{2\pi r}\cos\theta, \\ H_\theta &= \frac{lI}{2\pi r^2}\cos\theta, \end{aligned} \tag{139}$$

where l is the interaxial separation, r is the distance of a typical point of the field from the axis of the outer conductor, and θ is the remaining polar coordinate.

This field is impressed upon the inner surface of the outer conductor [33] and the resulting power loss equals, by equation (136), the real part of

$$\Phi = 2\pi a\eta\left(\frac{l}{2\pi a^2}\right)^2 I^2 = \frac{l^2}{2\pi a^3}\eta I^2, \tag{140}$$

where at high frequencies $\eta = \sqrt{i\omega\mu/g}$ is simply the intrinsic impedance of the outer conductor.[34] This loss increases the resistance of the outer tube by the amount,

$$\Delta R_a = \frac{l^2}{a^3}\sqrt{\frac{\mu f}{\pi g}}, \tag{141}$$

[32] For promixity effect in parallel wires external to each other, the reader is referred to the following papers: John R. Carson [1], C. Manneback [9], S. P. Mead [12].

[33] The radius of this surface is designated by a.

[34] At low frequencies η has to be replaced by the radial impedance looking into the shield.

in excess of the concentric resistance $R_a = (1/2a)\sqrt{\mu f/\pi g}$ given by (84). The relative increase is, therefore,

$$\frac{\Delta R_a}{R_a} = \frac{2l^2}{a^2}. \tag{142}$$

The magnetic field (139) is partially reflected from the outer tube, impressed upon the inner conductor, partially refracted into it and dissipated there. Using (110) and (111) we can show that the reflected field is

$$\begin{aligned} H_\theta &= \frac{lI}{2\pi a^2}\cos\theta, \\ E_z &= \frac{i\omega\mu lI}{2\pi a^2}\rho\cos\theta. \end{aligned} \tag{143}$$

This field converges to the axis of the outer conductor. In order to estimate its effect upon the inner conductor, it is convenient to replace it by an equivalent field converging toward the axis of the inner conductor. By properly changing the origin of the coordinate system this equivalent field can be shown to be

$$\begin{aligned} E_z &= -\frac{i\omega\mu lI}{2\pi a^2}(l + \rho\cos\varphi), \\ H_\varphi &= -\frac{lI}{2\pi a^2}\cos\varphi. \end{aligned} \tag{144}$$

Applying once more (138) (replacing there a by the radius b of the inner conductor), we find that the power loss due to this field is given by the real part of

$$\Phi = \frac{bl^2}{2\pi a^4}\ I^2, \tag{145}$$

so that the absolute increase in resistance of the inner conductor is

$$\Delta R_b = \frac{bl^2}{a^4}\sqrt{\frac{\mu f}{\pi g}} \tag{146}$$

which must be added to the concentric resistance of the inner conductor $R_b = (1/2b)\sqrt{\mu f/\pi g}$. The relative increase is therefore

$$\frac{\Delta R_b}{R_b} = 2\left(\frac{b}{a}\right)^2\left(\frac{l}{a}\right)^2. \tag{147}$$

It is unnecessary to carry the process further.

Considering the pair as a whole, the resistance when concentric is $R = R_a + R_b$, and the increase due to eccentricity is $\Delta R = \Delta R_a + \Delta R_b$ thus giving a percentage increase,

$$\frac{\Delta R}{R} = 2\left(\frac{b}{a}\right)\left(\frac{l}{a}\right)^2. \tag{148}$$

It is obvious that, so long as b and l are small compared with a, this percentage increase is very small.

From the well-known formulæ for the inductance and the capacity between parallel cylindrical conductors, we find that the characteristic impedance of a nearly coaxial pair is given in terms of the characteristic impedance of the coaxial pair by

$$Z = Z_0\left[1 - \frac{e^2k^2}{(k^2 - 1)\log k}\right], \tag{149}$$

where the "eccentricity" e is defined as the ratio of the interaxial separation to the inner radius of the outer conductor and k as the ratio of the inner radius of the outer conductor to the outer radius of the inner conductor. Combining (149) and (148) we have for the attenuation of the nearly coaxial pair:

$$\alpha = \alpha_0\left[1 + \frac{2e^2}{k} + \frac{e^2k^2}{(k^2 - 1)\log k}\right]. \tag{150}$$

References

Papers

1. John R. Carson, "Wave Propagation Over Parallel Wires: The Proximity Effect," *Phil. Mag.*, Vol. 41, Series 6, pp. 607–633, April, 1921.
2. John R. Carson and J. J. Gilbert, "Transmission Characteristics of the Submarine Cable," *Jour. Franklin Institute*, p. 705, December, 1921.
3. John R. Carson and J. J. Gilbert, "Transmission Characteristics of the Submarine Cable," *Bell Sys. Tech. Jour.*, pp. 88–115, July, 1922.
4. John R. Carson, "The Guided and Radiated Energy in Wire Transmission," *A.I.E.E. Jour.*, pp. 908–913, October, 1924.
5. John R. Carson, "Electromagnetic Theory and the Foundations of the Electric Circuit Theory," *Bell Sys. Tech. Jour.*, January, 1927.
6. S. Butterworth, "Eddy Current Losses in Cylindrical Conductors, with Special Applications to the Alternating Current Resistances of Short Coils," *Phil. Trans., Royal Soc. of London*, pp. 57–100, September, 1921.
7. H. B. Dwight, "Skin Effect and Proximity Effect in Tubular Conductors," *A.I.E.E. Jour.*, Vol. 41, pp. 203–209, March, 1922.
8. H. B. Dwight, "Skin Effect and Proximity Effect in Tubular Conductors," *A.I.E.E. Trans.*, Vol. 41, pp. 189–195, 1922.
9. C. Manneback, "An Integral Equation for Skin Effect in Parallel Conductors," *Jour. of Math. and Physics*, April, 1922.
10. H. B. Dwight, "A Precise Method of Calculation of Skin Effect in Isolated Tubes," *A.I.E.E. Jour.*, Vol. 42, pp. 827–831, August, 1923.
11. H. B. Dwight, "Proximity Effect in Wires and Thin Tubes," *A.I.E.E. Jour.*, Vol. 42, pp. 961–970, September, 1923; *Trans.*, Vol. 42, pp. 850–859, 1923.

12. Mrs. S. P. Mead, "Wave Propagation Over Parallel Tubular Conductors: The Alternating Current Resistance," *Bell Sys. Tech. Jour.*, pp. 327–338, April, 1925.
13. Chester Snow, "Alternating Current Distribution in Cylindrical Conductors," *Scientific Papers of the Bureau of Standards*, No. 509, 1925.
14. John R. Carson and Ray S. Hoyt, "Propagation of Periodic Currents over a System of Parallel Wires," *Bell Sys. Tech. Jour.*, pp. 495–545, July, 1927.
15. John R. Carson, "Rigorous and Approximate Theories of Electrical Transmission Along Wires," *Bell Sys. Tech. Jour.*, January, 1928.
16. John R. Carson, "Wire Transmission Theory," *Bell Sys. Tech. Jour.*, April, 1928.
17. A. Ermolaev, "Die Untersuchung des Skineffektes—Drahten mit Complexer Magnetischer Permeabilitate," *Archiv. f. Elektrotechnik*, Vol. 23, pp. 101–108, 1929.
18. Louis V. King, "Electromagnetic Shielding at Radio Frequencies," *Phil. Mag.*, Vol. 15, Series 7, pp. 201–223, February, 1933.
19. E. J. Sterba and C. B. Feldman, "Transmission Lines for Short-Wave Radio Systems," *Proc. I. R. E.*, July, 1932, and *Bell Sys. Tech. Jour.*, July, 1932.
20. H. Pleijel, "Beräkning af Motstånd och Självinduktion," Stockholm, K. L. Beckmans Boktryckeri, 1906.
21. H. Pleijel, "Electric and Magnetic Induction Disturbances in Parallel Conducting Systems," 1926, *Ingeniorsvetenskapsakademiens Handlingar* NR 49.
22. J. Fisher, "Die allseitige in zwei Kreiszylindrishen, konaxial geschichteten Stoffen bei axialer Richtung des Wechselstromes," Jahrbuch der drahtlosen Telegraphie und Telephonie, Band 40, 1932, pp. 207–214.

Books

J. Clerk Maxwell, "Electricity and Magnetism," Vols. 1 and 2.
O. Heaviside, "Electrical Papers."
Sir William Thomson, "Mathematical and Physical Papers."
Lord Rayleigh, "Scientific Papers."
Sir J. J. Thomson, "Recent Researches in Electricity and Magnetism."
A. Russell, "A Treatise on the Theory of Alternating Currents."
John R. Carson, "Electric Circuit Theory and the Operational Calculus."
R. W. Pohl, "Physical Principles of Electricity and Magnetism."
Max Abraham and R. Becker, "The Classical Theory of Electricity and Magnetism."
E. Jahnke and F. Emde, "Tables of Functions," B. G. Teubner, 1933.

Note: This list of references is by no means complete. Only the more recent papers dealing with some phase of the subject treated here are included.

Contemporary Advances in Physics, XXVIII
The Nucleus, Third Part *

By KARL K. DARROW

This article deals first with the newer knowledge of alpha-particle emission: that common and striking form of radioactivity, in which massive atom-nuclei disintegrate of themselves, emitting helium nuclei (alpha-particles) and also corpuscles of energy in the form of gamma-rays or high-frequency light. There follows a description of the contemporary picture of the atom-nucleus, in which this appears as a very small region of space containing various charged particles, surrounded by a potential-barrier; and the charged particles within, or those approaching from without, are by the doctrine of quantum mechanics sometimes capable of traversing the barrier even when they do not have sufficient energy to surmount it. The exponential law of radioactivity—to wit, the fact that the choice between disintegration and survival, for any nucleus at any moment, seems to be altogether a matter of pure chance—then appears not as a singularity of nuclei, but as a manifestation of the general principle of quantum mechanics: the principle that the underlying laws of nature are laws of probability. Moreover it is evident that transmutation of nuclei by impinging charged particles, instead of beginning suddenly at a high critical value of the energy of these particles, should increase very gradually and smoothly with increasing energy, and might be observed with energy-values so low as to be incomprehensible otherwise; and this agrees with experience.

Diversity of Energies in Alpha-Particle Emission

ON EVERYONE who studied radioactivity some twenty years ago, there was impressed a certain theorem, an attractively simple statement about the energy of alpha-particles: it was asserted that all of these which are emitted by a single radioactive substance come forth from their parent atoms with a single kinetic energy and a single speed. When beams of these corpuscles were defined by slits and deflected by fields for the purpose of measuring charge-to-mass ratio, nothing clearly contradicting this assertion was observed: the velocity-spectrum of the deflected corpuscles appeared to consist of a single line. In studying the progression of alpha-particles across dense matter, it was indeed observed that not all of those proceeding from a single substance had sensibly the same range. It is, however, to be expected that if two particles should start with equal energy into a sheet of (let us say) air or mica, they would usually traverse unequal distances before being stopped; for the stopping of either would

* This is the second and concluding section of "The Nucleus, Third Part," begun in the July, 1934 *Technical Journal*.

"The Nucleus, First Part" was published in the July, 1933 issue of the *Bell Sys. Tech. Jour.* (**12**, pp. 288–330), and "The Nucleus, Second Part" in the January, 1934 issue (**13**, pp. 102–158).

be brought about by its encounters with atoms and the electrons which atoms contain; and there would be statistical variations between the numbers of atoms and electrons which different particles would encounter after plunging into such a sheet. The probable effect of these variations can be computed; and it was shown at an early date that for at least some of the substances emitting alpha-rays—α-emitters—the diversity in ranges of the particles is no broader than should be expected. The curve of distribution-in-range of an α-ray beam often consists of a single peak or hump, and the shape and breadth of the hump are consistent with the assumption that it is due entirely to the "straggling" (the name applied to the statistical variations aforesaid) of particles all possessed initially of a single speed.

The vanishing of this beautiful but too-simple theorem from physics is due to experiments of three types. First, it was found that when all of the well-known α-particles of about 8.6 cm. range from ThC′ were completely intercepted by a stratum of matter of rather more than 8.6-cm. A.E. (air-equivalent [12]), and the detecting apparatus was adjusted to a sensitiveness much greater than would have been tolerable for the main beam, a very few particles—a few millionths of the number in the 8.6-cm. flock—were still coming through. Some of these had ranges as great as 11.5 cm., immensely greater than could be ascribed to straggling. These are the "long-range" alpha-particles, other examples of which have been discovered with RaC′ and (very lately) with AcC′. Next the colossal new magnet at Bellevue near Paris was employed by Rosenblum for deflecting α-particles and observing their velocity-spectrum, and the unprecedented dispersion and resolving-power (to employ optical terms) of this superb apparatus disclosed that for several α-emitters (the list now comprises eight) the spectrum consists of two or several lines instead of only one. The "groups" of alpha-particles to which these lines bear witness lie closer to one another in energy than the aforesaid long-range particles lie to the medium-range ones, wherefore they are often said to constitute the "fine-structure" of the alpha-rays; but it is probable that a more significant basis for distinction lies in the fact that the long-range corpuscles are relatively scanty, while the various lines of a fine-structure system are not so greatly unequal in intensity.

[12] I recall that while the range of an alpha-particle of given speed depends on the density and nature of the substance which it is traversing, the student is usually dispensed from taking account of this by the fact that the investigators nearly always state, not the actual thickness of the actual matter which they used, but the thickness of air at a standard pressure and temperature (usually 760 mm. Hg and 15° C.) which would have the same effect.

Another great magnet of a peculiar and original construction, developed at the Cavendish, was then applied both to spectra displaying fine structure and to the spectrum of RaC′, with notable success; while the technique of determining distribution-in-range curves has been improved to such an extent that it now almost rivals the magnets in its capacity of distinguishing separate groups in an alpha-ray beam. The theorem of the unique speed is therefore like so many another theorem of physics; it was valid so long as the delicacy of the experimental methods was not refined beyond a certain point, its validity ceased as soon as that point was passed.

To enter now into detail:

The *long-range particles* were discovered by observing scintillations, a method of singular delicacy and value, but having great disadvantages: all the observations being ocular, it is wearisome and taxing, not every eye is capable of it, and there is no record left behind except in the observer's memory or notes. Tracks of some of these particles were later photographed in the Wilson chamber, but it is a long research to procure even a few hundred of such photographs, and yet even a few hundred are not sufficiently many for plotting a really good distribution-in-range curve (the disagreements between the early work with scintillations and the subsequent work with Wilson chambers are rather serious). The best available curve is that which Rutherford, Ward and Lewis obtained with the method of the "differential ion-chamber," of which the principle is as follows:

When an alpha-particle (or, for that matter, a proton) traverses a sheet of matter, its ionizing power or ionization per-unit-length-of-path—we may take one mm. as a convenient unit of air-equivalent—varies in a characteristic way with the length of path which the particle has yet to traverse before being stopped completely. The ionization-curve at first is nearly horizontal, then rises to a pretty sharp maximum, then falls rapidly to zero.[13] Suppose now that the particle traverses a pair of shallow ionization-chambers, each containing a gas of which the thickness amounts to not more than a few mm. of air-equivalent (the same for both) and the two separated by a metal wall of negligible air-equivalent. Suppose further that the metal wall is both the negative electrode of the one chamber and the positive electrode of the other, and that it is connected to the electrometer or other detecting device. The charge which is perceived is then the difference between the ionizations in the two chambers. If these are traversed by a particle which is yet far from the end of its range, the difference

[13] The curve for protons is exhibited in Fig. 7 of "The Nucleus, Second Part," p. 124.

will be small and perhaps imperceptible; if by a particle which is approaching its maximum ionizing-power, the difference will be appreciable and of one sign; if by a particle which is coming to the end of its range, the difference will be considerable and of the opposite sign. So the differential chamber and its detecting device (in these experiments, an oscillograph connected through an amplifier, reacting appreciably to the passage of a single particle) are sensitive above all to particles which are nearing the ends of their ranges; and if a small number of such corpuscles be mingled with even a much greater number of faster charged particles—be they alpha-particles, be they protons, be they the fast electrons produced by gamma-rays—this circumstance, which would cripple any other method, will be almost without effect on it.[14] If the readings of the electrometer are plotted against the air-equivalent of the thickness of matter between the source (of alpha-particles) and the chamber, the resulting curve should not be much distorted from the ideal distribution-in-range curve.

The curve obtained in this way for the long-range particles of RaC′ exhibits a notable peak at range 9.0 cm.; to one side thereof a very much lower hump at smaller range (7.8); to the other side a wavy curve with four distinct maxima, which Rutherford and his colleagues deem to be the superposition not of four peaks only, but of seven. I show this portion (Fig. 5) to illustrate the analysis of such a curve for groups. Even the tallest of the peaks just mentioned is a mere molehill compared to the mountain which the principal group of RaC′ —the 6.9-cm. α-particles which were formerly the only ones known— would form if it could be plotted on the same sheet of paper; for the abundances of the 7.8-cm., 9.0-cm. and 6.9-cm. groups stand to one another as 1 : 44 : 2,000,000.

These, however, are not the latest words concerning the α-spectrum of RaC′. The energies of these groups might be deduced from their ranges, but for this purpose it is necessary to use an empirical curve of energy *vs.* range which at the time of the foregoing spectrum-analysis had been extended only up to range 8.6 cm. It was desirable to measure the energies of some of these groups directly, not only for their intrinsic interest but in order to carry onward that empirical energy-*vs.*-range relation. Recourse must therefore be had to a deflection-method.

Now in the usual form of magnet employed in deflection-experiments the field pervades the whole of the space between the solid disc-shaped faces of two pole-pieces. Were the pole-pieces to be so hollowed out

[14] Also a proton near the end of its range can be distinguished from an alpha-particle near the end of *its* range, on account of the difference in maximum ionization-per-unit-length ("The Nucleus, Second Part," pp. 124–125).

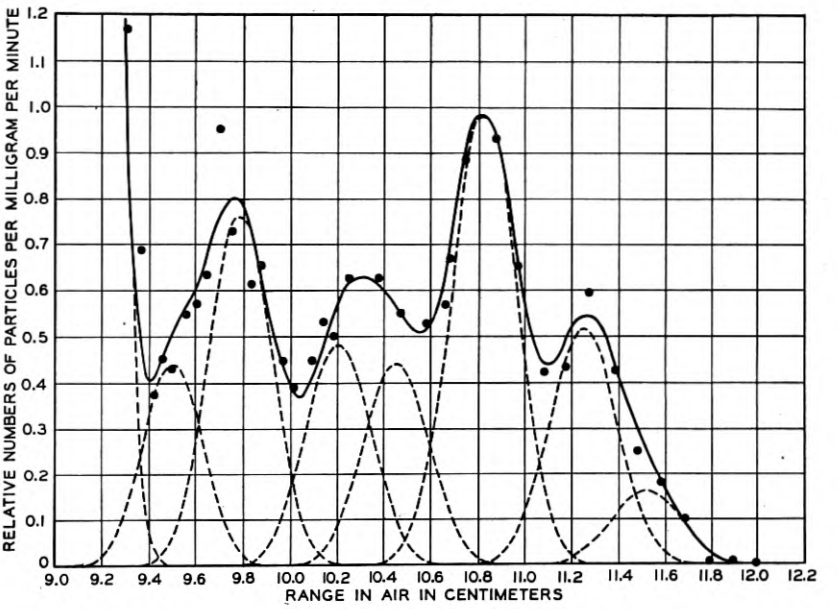


Fig. 5—Distribution-in-range of the long-range alpha-particles proceeding from RaC′, determined with the differential ionization-chamber (Rutherford Ward & Lewis; *Proc. Roy. Soc.*).

that their disc-shaped faces were reduced to narrow circular rings, there would be a great economy in magnetizable metal and a great reduction of weight and volume of the apparatus, as well as other advantages. This need not impair the availability of the magnet for analyzing a beam of α-particles, provided that the α-emitter can be located in the narrow annular space between the faces of the rings, and provided that the magnetization of the metal can be varied sufficiently widely. For then, for each group of α-particles there will be a certain value of the field-strength, whereby those particles which start out in directions nearly perpendicular to the field and tangent to the rings will be swept around in circular paths which are confined within that narrow annular space where alone the field exists. Somewhere in that space the detector should be placed; and the curve of its reading *vs.* field-strength H should show a peak for every group, and from the abscissa of the peak and the radius of the rings the energy of the group may readily be computed.

Such a magnet was built after Cockcroft's design at the Cavendish; the radius of its rings is 40 cm., they are 5 cm. broad and 1 cm. apart (these figures are the dimensions of the annular space), and the field-strength was adjustable up to 10,000 (later 12,000) gauss which was

sufficient for α-particles of energy up to and even beyond 10.6 MEV (millions of electron-volts) and range up to and even beyond 11.5 cm. Figure 6 exhibits the outward aspect, Fig. 7 the cross-section of this device (one sees how the armature is fully contained within the rings). The annular space and everything within it was evacuated (being walled in by the ring B seen in the figures); the detector—a simple ionization-chamber connected through a linear amplifier to an oscillograph—was set 180° around the annulus from the source. This device in the hands of Rutherford, Lewis and Bowden proved itself able to furnish even a better spectrum than the scheme of the differential ionization-chamber; all of the peaks indicated by the former curve were clearly separated, a hump which had suggested two groups was resolved into three maxima, and an extra group was discovered—*twelve* altogether! (Incidentally, the empirical energy-*vs.*-range curve of α-particles had previously been extended with the same device by Wynn-Williams and the rest, to energy = 10.6 MEV and range = 11.6 cm.)

The long-range spectrum of RaC′ is thus of no mean complexity. There will be occasion later for quoting its actual energy-values. Of the long-range spectrum of ThC′ there is relatively little to be said; evidently it has not been studied so intensively as the other, but it seems to be comparatively simple, for only two groups have been recognized. One of the groups of ThC′ has about the same range as the highest group of RaC′, so that between them they comprise the most energetic subatomic particles ever yet discovered (about 10.6 MEV) apart from those of the cosmic rays and those resulting from certain artificial transmutations. As for AcC′, it is one of the constituents of the mixture of radioactive bodies known as actinium active deposit, from which α-particles of a range of about 10 cm. have lately been observed in the Institut du Radium.

Thus far I have written as though RaC′ and ThC′ were isolable substances, of which one may obtain pure samples and analyze at leisure the α-rays thereof. The truth, however, is far otherwise; for the difficulties of making one radioactive substance practically free from others, serious in most cases, are utterly insuperable in these. Both RaC′ and ThC′ are so very ephemeral (their half-lives are too small to measure, and are guessed from the Geiger-Nuttall relation as 10^{-6} and 10^{-11} second respectively) that they can never be dissevered from their mother-elements RaC and ThC which are also α-emitters. Sometimes one finds the long-range particles designated as belonging to RaC or ThC, and indeed I have nowhere found stated any compelling reason for attributing them to the C′-elements rather than the C-

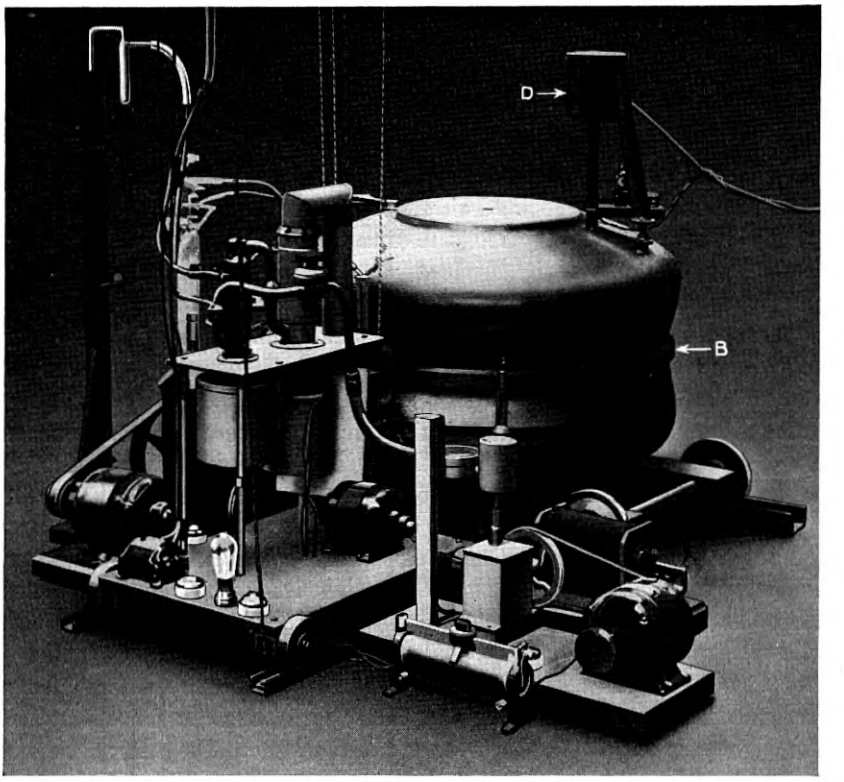


Fig. 6—Annular magnet employed for analysis of alpha-ray spectra. (After Rutherford, Wynn-Williams, Lewis & Bowden; *Proc. Roy. Soc.*).

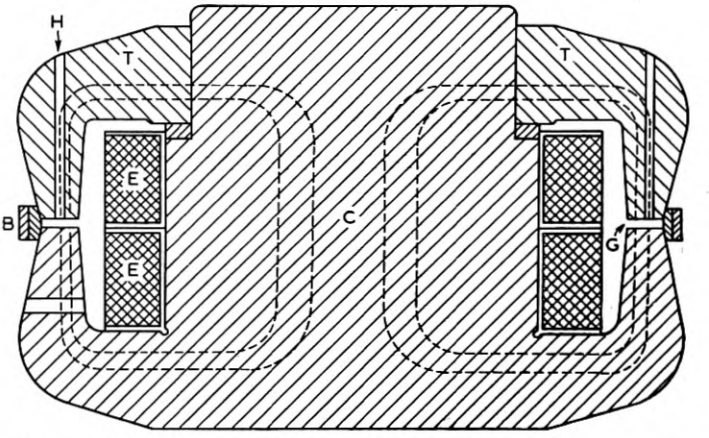


Fig. 7—Cross-section of the annular magnet used for analysis of alpha-ray spectra.

elements, apart from what is known about the correlated gamma-rays (see footnote 21).

Fine-structure was discovered, as I said before, with the great magnet of Bellevue. This has solid pole-pieces (75 cm. in diameter!) instead of rings; it is not necessary to adjust the field-strength step by step so as to bring group after group to a narrow detector; all the groups

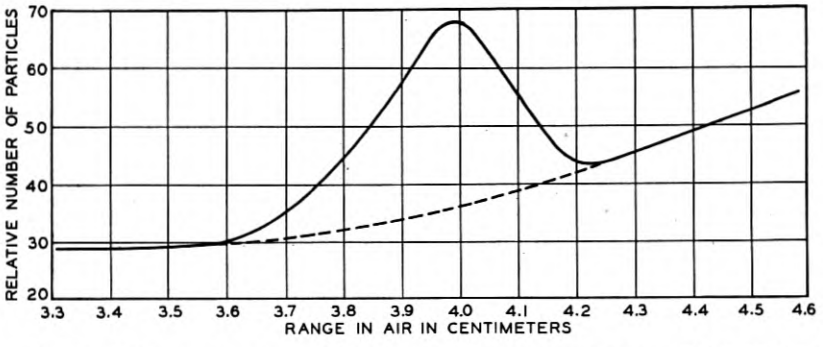


Fig. 8—Alpha-ray spectrum of RaC; peak observed with differential ionization-chamber, never before detected because of immensely greater number of particles in RaC′ peak just off the diagram to the right; asymmetry indicating fine-structure. (Rutherford, Ward & Wynn-Williams; *Proc. Roy. Soc.*).

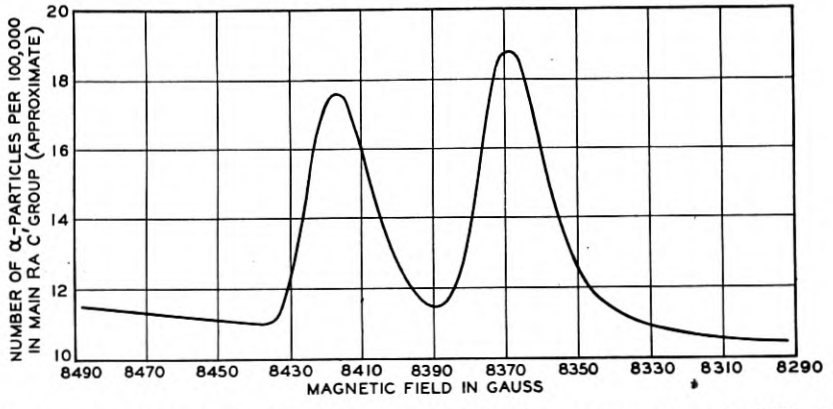


Fig. 9—Fine-structure of alpha-ray spectrum of RaC; as asymmetric peak of Fig. 8 resolved into two nearly equal peaks by annular magnet. (Rutherford, Wynn-Williams, Lewis & Bowden; *Proc. Roy. Soc.*).

of various speeds are deviated simultaneously in circular arcs each of its own particular radius, and simultaneously fall upon a photographic plate, producing what looks precisely like a line-spectrum in optics (Figs. 10, 11). The example in Fig. 10 relates to ThC, the earliest to

be analyzed; four lines only are visible upon the reproduction, but some plates after long exposure have shown as many as six.* The fine-structure of AcC consists of a pair of lines, which were detected as peaks in the distribution-in-range curve obtained at the Cavendish with a differential ionization-chamber. Instead of showing this curve I have chosen the corresponding curve for RaC, albeit it shows only a single hump (Fig. 8).[15] The unsymmetrical shape of this hump, how-

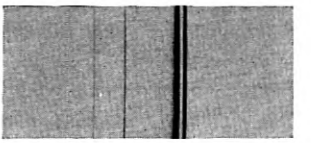


Fig. 10—Fine-structure of alpha-ray spectrum of ThC (not completely brought out in picture) obtained with Bellevue magnet. (S. Rosenblum.)

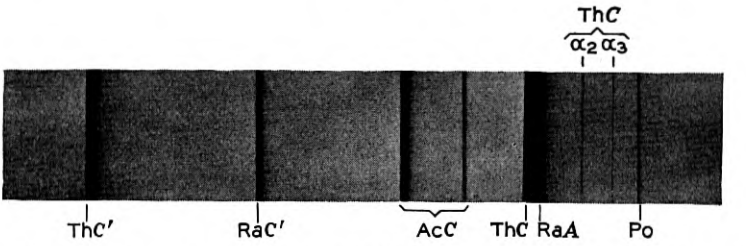


Fig. 11—Alpha-ray spectra of several elements (those of Po and AcC shifted slightly to the right with respect to the rest). (S. Rosenblum, *Origine des rayons gamma*, Hermann & Cie.).

ever, implies that it really consists of a pair of overlapping peaks; and so it does; for when the Cavendish magnet was applied to an α-ray beam from this element, the curve of detector-reading *vs.* field-strength displayed two equal peaks quite sharply separate (Fig. 9).

According to Rosenblum's latest census (February 2, 1934) there are now eight known examples of fine-structure: from the radium series, Ra (two groups), RaC (2); from the thorium series, RdTh (2), ThC (6); from the actinium series, RdAc (no fewer than eleven groups, the richest case of all!), AcX (3), An (3), AcC (2). According to Lewis and Wynn-Williams, there are (or were, in the spring of 1932)

* I am indebted to Dr. Rosenblum for a print from which Fig. 10 was made.

[15] This curve was the first to disclose the α-rays of RaC, previously known only by inference (though it was very compelling inference). Being of somewhat lesser range than the much more numerous (to be precise, 3000 times as numerous) α-particles emanating from the RaC′ atoms with which RaC is always inevitably mingled, they were completely hidden from observation by any method known before the use of the differential chamber and the powerful magnet.

at least five cases in which the distribution-in-range curve obtained with the differential chamber shows a single symmetrical peak suggesting only one group: from the radium series, Rn and RaA; from the thorium series, Tn and ThA; from the actinium series, AcA. Altogether there are twenty-three [16] known alpha-emitters, so that nearly half of the total remain to be investigated to this end. It may be significant that out of the four known alpha-emitters having odd atomic number, the high proportion of three at least is known to display fine-structure (the fourth, Pa, being as yet uninvestigated).

Interrelations of Alpha-Ray Spectra and Gamma-Ray Spectra

Evidently, if two atoms of the same radioactive substance were to emit alpha-particles of different speeds, there would be three obvious possibilities. The resultant nuclei might be different: in this case we should expect (though not with certainty) that they would be the starting-points of different radioactive series, and we should speak of "branching." The initial nuclei might have been different, in which case it would have been improper to speak of them as belonging to the same substance. Finally one at least of the two atoms might also emit gamma-ray photons, of energies complementary to those of the alpha-particles, in such a way that the total amount of energy released by the one atom would be the same as the total amount released by the other.

The first of these possibilities is not to be excluded *a priori* (since there are known cases of branching, though in them the alternative is between emission of an alpha-particle and emission of an electron) and neither is the second. The third, however, is the most agreeable, since if realized it allows us to believe that in the transformation of radium (to take one example) every radium nucleus is like every other before its change begins and every resulting (radon) nucleus is like every other after its change is over. Now alpha-ray emission and gamma-ray emission often occur together, which suggests that often the third possibility is the one which is realized; but this cannot be proved without measuring the energies or the wave-lengths of the gamma-rays.

The simplest cases are those in which the alpha-ray spectrum consists of two lines only. Here and always, there is an inconvenient complication at the start: when an alpha-particle is emitted, the residual nucleus recoils, and it is the sum of the kinetic energies of

[16] Not including Sm and other elements of atomic number lower than 81. The rest are depicted (together with the beta-ray emitters of atomic numbers 81 and greater) in Fig. 21.

the two (not that of the alpha-particle alone!) which must be taken into account.[17] Denote by U_1 and U_2 the values of this sum for the faster and for the slower alpha-particles. Does the gamma-ray spectrum then consist of a single line of which the photon-energy $h\nu$ is equal to $(U_1 - U_2)$?

In the case of AcC, the difference $(U_1 - U_2)$ is 0.35 or 0.36 MEV. There is an intense gamma-ray line proceeding from actinium active deposit (comprising AcC), and the energy of its photons is concordant. In the case of RaC, the difference $(U_1 - U_2)$ is only 0.04 MEV, and the search for so relatively soft a radiation of photons is difficult. In the case of Ra the conditions are more favorable, and here the history is worth retelling. Long before the earliest analysis of alpha-ray spectra, radium was known to emit feeble gamma-rays of photon-energy about 0.19 MEV. An estimate of their intensity was made in 1932 by Stahel; he concluded that the photons are less than one-tenth as numerous as the alpha-particles already known. Search was thereupon made by Rosenblum for fine-structure in the alpha-ray spectrum of radium. Two lines appeared on the plate after five minutes' exposure: they were due to groups proceeding one from radium and the other from its daughter-element radon. On plates exposed for hours there appeared yet another line. The values of U_1 and U_2 being computed for this and for the stronger radium group, the difference was found to be close to 0.185 MEV, with a sufficient latitude to be concordant with the estimate for the photons.

As the number of alpha-ray lines increases beyond two, the prospects rapidly become formidable; for a spectrum of n such lines suggests n possible states of the residual nucleus, and every one of these might "combine" (in the technical sense of the word) with every one below it in the energy-scale, making a total of $n(n-1)/2$ gamma-ray lines to be expected. Even so, anyone acquainted only with optical spectra might think it no difficult matter to photograph (say) the gamma-ray spectrum of ThC, and see whether it consists in just 15 lines in just the right places to correspond with the six alpha-particle groups. But one does not photograph gamma-ray spectra—one photographs the beta-ray spectra of the electrons ejected by the gamma-rays from atoms, and tries to deduce the photon-energies $h\nu$ from the electron-energies.[18] The atoms may be those of the radioactive substance

[17] By multiplying the kinetic energy of the alpha-particle by the factor $(1 + m/M)$, where m stands for the mass of the alpha-particle and M for that of the recoiling nucleus. This point was overlooked by a number of people before it was noticed by Feather.

[18] I have dealt with this procedure at length in the article "Radioactivity," No. XII of this series (*Bell Sys. Tech. Jour.*, 6, 55–99, 1927).

itself (either the very ones which are emitting the gamma-rays, in which case the rays are said to undergo "internal conversion," or their neighbours) or they may be those of other elements mixed with the radioactive substances, or those of nearby solids or gases on which the gamma-rays fall. Each gamma-ray line is responsible for several different beta-ray lines, a circumstance which makes the analysis more difficult at the beginning though it makes the inference more reliable in the end. There may be gamma-rays having nothing to do with alpha-particle emission, and there may be gamma-rays from several different radioactive substances inextricably mixed up together, so that the problem of analyzing the spectrum of one transformation is preceded (or, more truly, accompanied) by that of distinguishing it from the intermingled spectra of others.[19] The experimental errors in the estimates of U-values and $h\nu$-values may be so large that apparent agreements are actually unreliable. Altogether, the comparison of a rich alpha-ray spectrum with a rich gamma-ray spectrum is an exceedingly intricate business, the outcome of which is not to be summarized in a few sentences. To give a mere notion of the sort of conclusion which is reached, I quote some lines from Rutherford, Lewis and Bowden, in their comparison of the thirteen-line alpha-ray spectrum and the very rich gamma-ray spectrum of RaC′:

"When we consider in broad terms the data which have been presented, there can be no doubt that there is a high correlation between the alpha-particle levels which have been observed and the emission of gamma-rays. In more important cases the numerical agreement is well within the experimental error of measurement, while the relation between the intensity of the alpha-ray groups and the gamma-rays associated with them is of the right order of magnitude to be expected on general theoretical grounds. In other cases the agreement is very uncertain, and more definite information on the gamma-rays is required to make the deductions trustworthy. It is unfortunate that we have been unable to detect the alpha-particle groups corresponding to certain postulated levels [*i.e.* postulated from the classification of the gamma-ray lines] . . . ."

Thus it appears that there are excellent agreements between $h\nu$-values and $(U_i - U_j)$ values, and yet nothing approaching a perfect one-to-one-correspondence. Nevertheless, the general programme is fixed: to assume that each nucleus possesses a system of stationary

[19] It is interesting to notice that after the $h\nu$-values of certain gamma-rays emitted from mixtures of ThC and ThC″ had been found to agree with values of $(U_i - U_j)$ taken from the alpha-ray spectrum of ThC, these gamma-rays were proved to proceed from ThC in its transformation into ThC″, whereas till then they had been supposed to proceed from ThC″ in its transformation into ThD (Meitner & Philipp, Ellis).

states and energy-levels, to assume further that $h\nu$-values and $(U_i - U_j)$-values are alike the differences between these energy-levels, and to ascribe apparent defects of correlation to special circumstances by virtue of which certain gamma-rays and certain alpha-rays are too feeble to be detected. Should these ideas prove untenable, we shall probably have to suppose that the nucleus is even more different from the extra-nuclear world than we have hitherto admitted.

Now arises the important question: when alpha-particles and photons both are emitted in the course of the complete transformation of one nucleus into another, which comes first? Despite the immeasurable shortness of the times which are involved, this is in principle an answerable question. For as I have mentioned already, gamma-rays are detected and their photon-energies are measured by examining the spectrum of the electrons which they eject from the orbital electron-layers of atoms, chiefly from the layers surrounding those very nuclei whence the photons themselves proceed. Now if the photons come before the alpha-particles, say for example in the transformation of ThC into ThC″, these electrons will come from the electron-layers of ThC atoms; in the contrary case, from the layers of ThC″. It is possible to distinguish from which they do come, even when the energy of the photons is not independently known and must itself be derived from the same data.[20]

The classical and crucial experiment of this type was performed about ten years ago by Meitner, and it proved that the gamma-rays emitted during the transformation of RdAc into AcX and during that of AcX into An spring forth *after* the alpha-particle has departed and the nucleus has become that of the daughter-element. These are two of the cases in which the alpha-ray spectrum exhibits fine-structure; and it is now generally supposed that the rule extends to all such cases. The stationary states or energy-levels deduced from the $(U_i - U_j)$-values and the $h\nu$-values then would pertain to the "final" or daughter nucleus. In the instances where all the alpha-particle groups except the main one are designated "long-range groups"—RaC′ and ThC′ (the quotation above from Rutherford, Lewis and Bowden refers to the former of these)—Gamow argues that the gamma-rays are emitted before the alpha-particles; the energy-levels deduced from the alpha-ray and the gamma-ray spectra would then pertain to

[20] See the previously-cited article "Radioactivity," pp. 94–96. I mention in passing that sometimes the "internal conversion" of photons whereby electrons are ejected is apparently so much the rule, that no appreciable fraction of the gamma-rays of some particular energy (or energies) escape from the atoms at all; in which cases it becomes expedient to speak not of gamma-rays at all, but of an immediate transfer of energy from the nucleus to the orbital electrons (a policy which may be applied to all cases of internal conversion).

the "initial" or mother nucleus. It is not clear from the literature whether this hypothesis has been fully tested in the manner of Meitner's tests aforesaid, but presumably it was adopted in calculating the $h\nu$-values from the electron-energies, so that the agreements between $h\nu$ and $(U_i - U_j)$ support it.[21]

The search for interrelations among the energy-levels, the different $h\nu$-values and the different U-values belonging to individual transformations has of course already begun. Rutherford and Ellis find that the frequencies of many of the lines in the gamma-ray spectrum of RaC′ can be fitted by assigning various integer values to p and q and constant values to E_1 and E_2 in the formula $pE_1 + qE_2$; while H. A. Wilson finds that if the U-values or the $h\nu$-values are added together in pairs, an amazing number of the pairs are equal to integer multiples (the integer multipliers ranging from 16 to 54) of the amount 0.385 MEV—this even if the two members of a pair are taken from different spectra!

The Quantum-Mechanical Theory and the Crater Model of the Nucleus[22]

Anyone who is acquainted with the contemporary atom-model in its present or in its earlier stages, with its congeries of charged particles revolving in or jumping between definitely-prescribed and quantized orbits, governed by attractions and repulsions both classical and unimaginable—any such person will probably be looking for a nucleus-model of the same variety but built on a very much smaller scale,

[21] I learn by letter from Dr. Ellis that in the case of RaC′, some at least of the gamma-rays which agree with the $(U_i$-$U_j)$-values of the long-range alpha-particles are definitely proved in this fashion to proceed from nuclei of atomic number 84 (that of RaC′) as distinguished from 83, 82 or 81; the proof is especially strong for the most intense gamma-ray, of photon-energy 0.607 MEV. Perhaps this is the most powerful evidence that the long-range particles come from RaC′ rather than RaC.

The half-periods of RaC′ and ThC′ are exceedingly short, 10^{-6} sec. and 10^{-11} sec. respectively; had it been otherwise, objection might be made to Gamow's contention on the ground that atoms in states, from which they are liable to depart by emitting radiation, generally do depart from those states and emit that radiation within a period of the order of 10^{-7} or 10^{-8} sec. There is no *a priori* certainty that this principle applies to nuclei, but if it does it may suffice to explain why the long-range particles are observed only from these very short-lived nuclei, why they are so scanty even in these cases (nearly always the photon *is* emitted before the alpha-particle gets ready to leave, so that the latter nearly always leaves with low energy instead of high), and why the fine-structure of other alpha-ray spectra is related to energy-levels of the final instead of the initial nucleus (Gamow). Incidentally it strengthens the case for ascribing the long-range particles to the C′-products instead of the C-products.

[22] Quantum mechanics was first applied to the nucleus-model here to be described, independently and almost simultaneously, by Gurney and Condon and by Gamow. Rather than by all three names together, it seems preferable to denote this model thus interpreted by a neutral descriptive term, such as "crater model" (an allusion to the aspect of the graph obtained when the curve of Fig. 12 is coupled with its mirror-image in the yz-plane).

with protons and possibly neutrons figuring among the revolving particles. He will be looking too far into the future, and will be disappointed with the present. The present nucleus-model consists of little more than a single curve—a curve which, moreover, relates only to the fringe of the nucleus and to the region surrounding it, and for want of knowledge is not extended into the central region or nucleus proper where the constituent particles must be. The theory which it serves is a theory not of the nucleus as a stable system of corpuscles, but of the escape of some from among these corpuscles and the entry of new ones—a theory professing to deal only with the entry and the escape, not at all with the events succeeding the one or preceding the other.

The curve purports to portray the electrostatic potential, as function of r the distance measured from the centre of the nucleus, from $r = \infty$ inward to a minimum distance which is indeed very small even in the atomic scale—10^{-13} cm. or less—but still definitely not zero, since the components of the nucleus must be presumed to be normally at distances yet smaller. When it is plotted as in Fig. 12, its ordinate

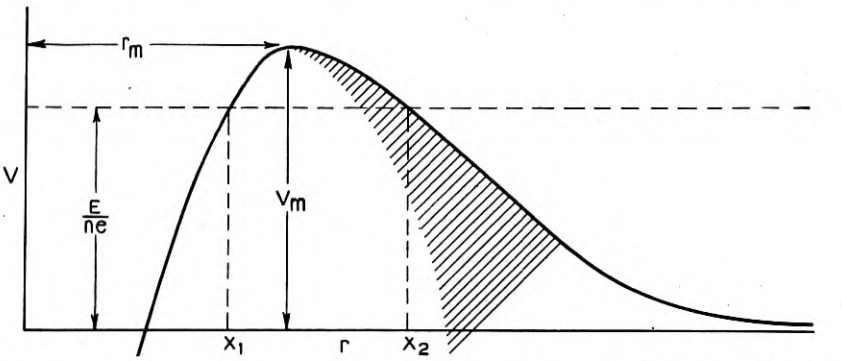


Fig. 12—Nuclear potential-curve postulated for explaining transmutation (without allowance for resonance) and radioactivity.

at any r is a measure of the amount of kinetic energy which a positively-charged particle approaching the nucleus must sacrifice—i.e. which must be converted into potential energy—in order to come from infinity to r. Traced from infinity inward, the curve must follow at first the function *const.*/r, corresponding to the inverse-square law of force; for it is known, both from experiments on alpha-particle scattering (which supplied the foundation for the contemporary atom-model) and from the successes of the theory of atomic spectra, that beyond a certain distance a nucleus is surrounded by an inverse-

square force-field. This is the obstacle, or at any rate, a part of the obstacle, which an oncoming proton or deuton or alpha-particle must overcome in order to reach the nucleus and achieve transmutation. One may picture it as a hill, up which the ball must roll to reach the castle at the top—and *down* which the ball will roll if it starts from the top, shooting outward towards infinity as the fast-flying alpha-particle.

Now to assume the inverse-square force as prevailing all the way inward to $r = 0$ would be to postulate a point-nucleus without room for parts or structure, surrounded by a hill of infinite height which no approaching positive particle could climb; all of which is inadmissible. Departure from the inverse-square law is actually shown by some experiments on scattering of alpha-particles which pass very close to the nucleus, and these indications are to be heeded in tracing that part of the curve of Fig. 12 which lies to the right of the maximum; but the maximum itself and the sharp descent to its left are dictated by no such observations, and to postulate them is to make the theory which is now to find its employment and its test. That there should be such a maximum and such a descent is of course the most natural supposition to make. If there are several particles of positive charge which stay for a finite time within the nucleus, there must be something which restrains them from flying away. This something must either be an agency of a type as yet unknown, or else be described by a potential-curve with a maximum at what we may henceforward call the boundary of the nucleus; and the latter assumption is to be preferred till proved unusable.

Applying classical ideas to this "model" (if the word be not considered too presumptuous) of a nucleus, one is led at once to two predictions, which may be sharply formulated if we adopt symbols such as V_m and r_m for the two parameters indicated on Fig. 12, *viz.* the "height of the potential-barrier" and the "radius of the potential-barrier" as they are commonly called, the latter being also called the "radius of the nucleus." These are:

1. If the nucleus emits a particle of positive charge $+ 2e$, the kinetic energy with which this particle is endowed when it completes its escape cannot be less than $2eV_m$; consequently, when it is observed that atoms of a certain element emit alpha-particles with kinetic energy K_0, the height of the potential-barrier for that element cannot surpass $K_0/2e$; consequently, when the force-field about the nuclei of such atoms is explored by the classical method of studying the scattering of alpha-particles projected against a sheet of that element, it must be found that the region of repulsive force, and *a fortiori* the

inverse-square field, do not extend far enough inward for the integral to surpass the value $K_0/2e$.

2. When particles of charge ne (n = any integer) are projected at a sheet of any specific element, they cannot enter the nuclei at all unless their kinetic energy exceeds the critical value neV_m; and if the curve of number-entering-nuclei *vs.* kinetic-energy can in any way be deduced from any experiments, it should rise fairly sharply from the axis of kinetic energy at this critical abscissa.

(It will have been noticed that I expressed both of these predictions as though the escape or the entry of a particle made no difference to the height of the potential-barrier, which is the universal practice. This is obviously too crude an assumption; the error in it must be graver the smaller the atomic number of the element, therefore graver in theorizing about the transmutation of light elements than in theorizing about the radioactivity of heavy ones; it must be rectified in future.)

The former of these predictions can be sharply and unquestionably tested; and it proves to be wrong. Uranium I. emits alpha-particles of kinetic energy K_0 equal to 4 MEV; but Rutherford suspected from scattering-experiments on other heavy elements, and subsequently proved by such experiments upon uranium itself, that the inverse-square force-field extends so far inwards as to involve a height of potential barrier at least twice as great as $K_0/2e$; so that an emerging alpha-particle should possess at least 8 MEV of kinetic energy derived from coasting down the hill, and even this is merely a lower limit to the estimate, since the hill may be higher and the particle might come with some excess of energy over its brow!

The second prediction is not so readily tested. If all of the charged particles (protons or deutons or alpha-particles, say) projected at the postulated sheet of matter were directed straight towards the centres of nuclei, and arrived at the potential-hills without suffering any prior loss of energy elsewhere, the fraction entering through the potential barriers would rise suddenly from zero to unity as the kinetic energy K of the particles was raised to neV_m, and any phenomenon depending solely upon entry would make its advent suddenly if at all. Unfortunately this does not occur in any experiment now possible or likely ever to become possible. If the sheet of matter is a monatomic layer, most of the oncoming particles will be going towards the gaps between the nuclei, and the initial directions of the rest will be pointed towards all parts of the cross-sections of the nuclei, only an infinitesimal fraction going straight toward the centres. Designate by p the perpendicular distance from a centre to the line-of-initial-motion of

an oncoming particle; it is evident that the minimum kinetic energy permitting of entry will increase with p, starting from neV_m and rising to infinity as p rises from 0 to r_m. The relation between fraction-of-particles-entering-nuclei—call it P_e—and kinetic energy K could be calculated, given specific assumptions about the values of V_m and r_m and the trend of the potential-curve. Without undertaking the calculation, it is easy to see that the vertical rise of what I will hereafter call the "ideal" curve—the curve of probability-of-entry-at-central-impact *vs.* K—will be distorted into a bending slope, starting, however, at the same critical abscissa neV_m. If the sheet of matter is a thick layer, there will of course be a much greater fraction of the impinging particles of which the initial paths point straight toward some nucleus or other, but the fraction achieving entry will not be raised in the same ratio, for the particles going toward nuclei embedded deep in the layer will lose some or the whole of their velocity in passing through the intervening matter.[23] This also will contribute to converting the vertical rise into a gradual bend. Still it does not seem possible that if the ideal curve had such a shape, the experimental ones could rise with so extreme a gradualness as does the one of Fig. 17 or those of Figs. 16 and 17 in the Second Part; for these suggest no sudden beginning at all, but rather they have the characteristic aspect of curves asymptotic to the axis of abscissas, as if their apparent starting-points could be pushed indefinitely closer to the origin by pushing up indefinitely the sensitiveness of the apparatus. Neither does it seem possible that V_m can be so low as their starting-points imply.

There is, however, another difficulty: these curves refer not directly to P_e, but to number of transmutations, or to be precise (for precision is essential in these matters) to the number of particles producing transmutations involving the ejection of fragments having certain ranges. Call this number P_t. It is easiest to conduct the argument as though P_t were proportional to P_e—as if an observable transmutation could result only from the entry of a particle through the potential-barrier of a nucleus, and as if the number of transmutations of any special type were strictly proportional to the number of entries, the factor of proportionality being independent of K. Yet few assumptions are less plausible. It is far more reasonable to suppose that the probability of a particle bringing about a transmutation when it enters a nucleus is not invariably unity, but is instead some function $f_1(K)$. It is reasonable also to suppose that a particle passing close to the potential-barrier but not traversing it may yet be able to touch off an internal explosion or eruption leading to a transmutation. Denote

[23] Contrast the two curves of Fig. 17.

by $f_2(K)(1 - P_e(K))$ the number of cases in which this happens. The least which we can take for granted is some general relation of the form,

$$P_t = f_1(K) \cdot P_e(K) + f_2(K)(1 - P_e(K)), \tag{20}$$

and the variations of f_1 and f_2 may contribute still further to blotting out all signs of the hypothetical vertical rise in the ideal curve. Moreover, f_2 might be appreciable at values of K smaller than neV_m, thus blotting out every sign of the critical energy-value at which entry commences.

Thus with regard to the second prediction, the situation is this: the experimental curves of number-of-observed-transmutations *vs.* kinetic-energy-of-impinging-particles rise so smoothly and so gradually from the axis as to give not the slightest support to the idea that entry into the nucleus commences suddenly at a critical value of K; moreover, transmutation commences to be appreciable—for several elements, at least—when K is still so small that K/ne is only a small fraction of the least value which can reasonably [24] be ascribed to V_m, in view of what we know from alpha-particle-scattering about the circumnuclear fields of these or similar elements. This again might be due to the hypothetical effect to which the term $f_2(K)$ in the equation alludes, but it seems far too prominent for that! With it is to be linked the fact that alpha-particles emerge from nuclei with kinetic energy less than $2eV_m$. The potential-hill seems not to be so high either for entering or for emerging particles, as it is for those which only skirt its slopes!

Now if in theorizing about potential-hills and particles we substitute quantum mechanics for classical mechanics, these phenomena cease to be things contrary to expectation, and become instead the very things to be expected.

This is one of the situations—regrettably frequent in the present-day theoretical physics—where neither pictures nor words are adequate. The nearest description which can be made with words is probably somewhat as follows: We set out to ascertain whether a particle of charge ne and kinetic energy K, coming from infinity straight toward the nucleus (I simplify the problem as much as possible) will surmount the potential-hill of height V_m. Were we to conceive it as a particle conforming to classical mechanics, we should arrive at the answers: yes, if $K \geqq neV_m$—no, if $K < neV_m$. But we are to turn away from

[24] It is true that the elements of which the circumnuclear fields have been most carefully explored by alpha-particles are not in general the same as those for which transmutability has been observed down to very low values of K; but boron and carbon figure on both the lists, Riezler having studied the scattering of alpha-particles by these (*Proc. Roy. Soc.*, **134**, 154–170, 1932).

the particle for awhile, and to conceive a train of waves advancing from infinity towards the nucleus. The phase-speed and the frequency of this wave-train are prescribed by definite rules making them dependent upon E, and the train is governed by a prescribed wave-equation in which figures the function $V(r)$ of Fig. 12. On solving this equation in the prescribed fashion we find that it requires the wave-train to continue (though reduced in amplitude) past the top of the hill if K is greater than neV_m. This is partially satisfactory, for the particle when it is reintroduced is to be associated with the waves, and everything would be spoiled if the particle could go where the waves cannot. But also, the equation requires the wave-train to continue past the top of the hill when K is less than neV_m. True, it does not wholly pass; there is a reflected as well as a transmitted beam, and the ratio of reflected to incident amplitude goes very rapidly up towards unity and the ratio of transmitted to incident amplitude goes very rapidly down toward zero as K drops downward from the value neV_m. All the same there *is* this wave-train beyond the hill with an amplitude greater than zero; and the association of particles with waves is apparently spoiled, for the waves can go where the particle cannot.

At this point, however, it is the rule of theoretical physicists to give the precedence to the waves, and declare that *where the waves go there the particle must go also, whether it can* (by classical mechanics) *or cannot.* Since some of the waves are beyond the hill, the particle also must be able to traverse the hill, even though its kinetic energy is insufficient for it to climb to the top. But since the waves beyond the hill have a smaller amplitude than those coming up from infinity, it is not certain that the particle will pass through, but merely possible. The chance or probability of its passing through is determined chiefly (not fully) by the ratio of the squared-amplitudes of the waves on the two sides of the hill, and this is what must be computed by quantum-mechanics. How the particle gets over or through the hill—where and what it is and how it is moving while it is getting through—these are questions which the theorist usually declares to be unanswerable in principle, and having so declared, he does not attempt to visualize this part of the process.

Into Fig. 12 the diagonal lines have been introduced in a crude attempt to make graphic as much as possible of the theory. The length of the sloping line drawn from any point P of the curve is meant as a sort of inverse suggestion of the chance which a particle of charge $+ ne$ has of entering the nucleus if its energy E is equal to ne times the ordinate of P: the longer the line, the less the chance of

entry! (This energy E will be the same as the initial kinetic energy K already so often mentioned, which the particle has before it starts to climb the hill.) Perhaps it is not too fanciful to think of these lines as the posts of a fence standing up vertically from the curve, the varying height of which is a rough indication of the varying difficulties which particles of various energies have in getting through.

To predict successfully how the height of this metaphorical fence, the probability of transmission or of penetration, varies with V or E would be a magnificent triumph of nuclear theory, but it is vain to hope for such a success in the immediate future. Almost certainly the top of the fence curves much more rapidly upward than the drawing suggests, and also there is good reason to think that there may be gaps in the fence somewhere like those depicted in Fig. 13, where the

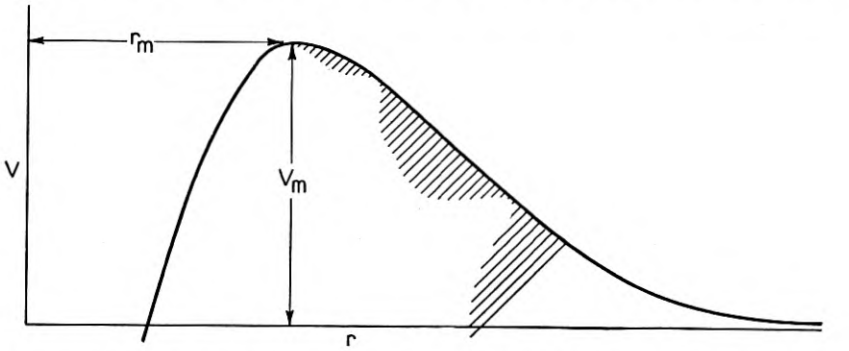


Fig. 13—Nuclear potential-curve postulated for explaining transmutation, with allowance for "resonance."

probability of penetration rises to values remarkably near to unity. Such at least are the features of certain one-dimensional potential-fields (do not forget that Figs. 12 and 13 refer to three-dimensional potential-fields having spherical symmetry!) which have isolated potential-hills or hills-adjoined-by-valleys.

Three of these cases are displayed in Figs. 14, 15, 16. Take the first for definiteness. One plunges *in medias res* by writing down at once Schroedinger's wave-equation:

$$d^2\Psi/dx^2 + (8\pi^2 m/h^2)[E - neV(x)]\Psi = 0, \qquad (21)$$

in which $V(x)$ stands for the potential-function exhibited in the figure,[25] while the meanings of ne, m and E have probably already been

[25] I deviate from the otherwise-universal usage of employing V for the potential energy of the particle, for the reason that the latter depends on the charge of the particle, while the potential-function is supposed (no doubt inaccurately) not to depend on it.

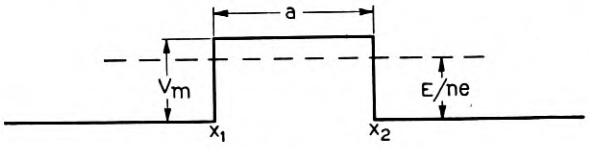


Fig. 14—Illustrating an artificial case of a potential-curve with a single square-topped potential-hill.

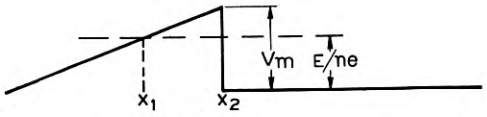


Fig. 15—Illustrating an artificial case of a potential-curve with a pointed hill.

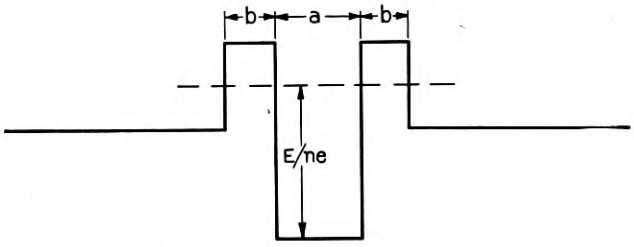


Fig. 16—Illustrating an artificial case of a potential-curve with a valley between two hills.

guessed by the reader—they are constants to which any values may be assigned, and the eventual result of the mathematical operations is going to be taken as referring to a stream of particles of charge *ne*, mass *m* and energy *E*.

The problem is stated as that of finding a solution of (21) for whatever value is chosen for *E*—a solution everywhere single-valued, bounded, continuous, and possessed of a continuous first derivative, such being the general requirement in quantum mechanics. Not, however, *any* solution possessing these qualities, but a solution apt to the physical situation. On the right of the hill, it must specify a wave-train (I) going from right to left; for we are interested in the adventures of particles coming from the right toward the hill. But on the right of the hill, it must also be capable of specifying a wave-train (II) going from left to right, for some or all of the particles may be reflected from the hillside. On the left of the hill it must be capable of specifying a wave-train (III) going from right to left, for some or all of the particles may traverse the hill and continue on their way.

So far as the region to the right of the hill ($x > x_2$) is concerned, a solution having all of these qualities is the following:

$$\Psi = A_1 e^{-ikx\sqrt{E}} + A_2 e^{+ikx\sqrt{E}}, \qquad k = \sqrt{\frac{8\pi^2 m}{h^2}}, \tag{22}$$

in which the two terms stand for wave-trains I and II, and A_1 and A_2 are adjustable constants. So far as the region to the left of the hill ($x < x_1$) is concerned, a solution having all of the required qualities is the following:

$$\Psi = C_1 e^{-ikx\sqrt{E}}. \tag{23}$$

It stands for wave-train III and C_1 is an adjustable constant. As I have already said, our "intuition" based on notions of what particles should do, expects C_1 to vanish and A_2 to become equal to A_1 when E is less than neV_m, but the solution of (21) does not consent to these limitations. Our intuition also expects that when E is less than neV_m nothing will happen in the region comprised within the hill ($x_1 < x < x_2$), but here again the solution of (21) does not conform with it. For in this region comprised within the hill, the solution must take the form:

$$\Psi = B_1 e^{-kx\sqrt{neV_m - E}} + B_2 e^{+kx\sqrt{neV_m - E}}, \tag{24}$$

which looks at first glance like (22) but is essentially different, since the exponents are real and not imaginary, and the terms do not represent progressive waves. The five coefficients—A_1, A_2, B_1, B_2, C_1—must now be mutually adjusted so that at the sides of the hill ($x = x_1$ and $x = x_2$) the expressions (22) and (23) and (24) flow smoothly each into the next, with no discontinuity either of ordinate or of slope. This imposes four conditions on the five coefficients, and therefore fixes the relative values of all of them—in other words, determines them completely except for a common arbitrary factor which corresponds to the intensity of the incident beam, and is irrelevant to the course of the argument.

In particular, this requirement of continuity imposed by the fundamental principles of quantum mechanics upon the acceptable solution of (21) fixes the ratio of the amplitudes C_1 and A_1 of "transmitted" and "incident" wave-train. From Gurney and Condon I quote an approximate formula [26] for the ratio of the squares of these amplitudes, denoting them on the left by the customary symbols:

$$\begin{aligned} \frac{(\Psi\Psi^*)_{\text{trans.}}}{(\Psi\Psi^*)_{\text{inc.}}} &= \phi(E) \exp.\,[-(4\pi a/h)\sqrt{2m(NeV_m - E)}], \\ \phi(E) &= 16(E/neV_m)(1 - E/neV_m). \end{aligned} \tag{25}$$

[26] Exact formula given by E. U. Condon, *Reviews of Modern Physics*, **3**, 57 (1931).

This expression does not vanish suddenly as soon as E drops below neV_m, but falls away continuously—and very rapidly, it must be admitted, owing to the exponential factor—as E diminishes from neV_m on downwards. Its value depends on a, the breadth of the hill (Fig. 14) in such a way that the broader or thicker the hill of given height the less the amplitude of the transmitted waves: the thicker the hill, the more nearly it comes to fulfilling the classical quality of being a perfect obstacle to particles having insufficient energy to climb it!

I rewrite (25) in the equivalent form,

$$\frac{(\Psi\Psi^*)_{\text{trans.}}}{(\Psi\Psi^*)_{\text{inc.}}} = \phi(E) \exp.\left[-(4\pi/h)\int_{x_1}^{x_2}\sqrt{2m(NeV_m - E)}dx, \quad (26)\right.$$

the integral in the exponent being taken "through the hill" from x_1 to x_2. This form is generalizable. Take the case of Fig. 14: the ratio of the squared-amplitudes of transmitted and incident wave-trains is given, according to Fowler and Nordheim, by an expression which is of the type (26), except that $\phi(E)$ is a somewhat different function (it is $4[(E/NeV_m)(1 - E/NeV_m)]^{1/2}$). The distance from x_1 to x_2, over which the integration is carried, obviously depends on E in this and every other case but the particular one of Fig. 14. Take finally the general case of a rounded hump, such as appears in Fig. 12. According to Gamow, a formula of type (26) is approximately—not exactly—valid for every such case, $\phi(E)$ being given by him as simply the number 4 when the hill descends to the same level on both sides as in Fig. 14; while in the general case where the potential-curve approaches different asymptotes at $-\infty$ and $+\infty$—say zero at the latter, V_r at the former—the factor $\phi(E)$ assumes the form $4[E/(E - neV_r)]^{1/2}$. Now E was the kinetic energy of the particles at infinity in the direction whence they come, and $(E - neV_r)$ will be the kinetic energy of the particles at infinity in the direction whither they are going. We have been denoting the first of these quantities by K; denote the second by K_r, and the corresponding velocities by v and v_r. Then $\phi(E)$ can be written as $4v/v_r$.

The question must now be answered: what is the actual relation between the ratio $(\Psi\Psi^*)_{\text{trans.}}/(\Psi\Psi^*)_{\text{inc.}}$, and the probability that a particle will traverse the hill? In associating waves with corpuscles, it is the rule to postulate that the square of the amplitude of the waves at any point is proportional to the number-per-unit-volume of corpuscles in the vicinity of that point. If one prefers to think of a single particle instead of a great multitude, one may say that the square of the amplitude of the waves at any point is proportional to the proba-

bility of the particle being at that point. Let us hold, however, to the picture of a dense stream of corpuscles approaching the potential-hill—say that of Fig. 14—from the right, and a much weaker stream receding from it on the left. If there are s times as many corpuscles per-unit-volume on the right as on the left, then there cannot be a steady flow unless only one out of s incident particles traverses the hill. The reciprocal of s is the fraction of particles getting through the hill, or the probability of a single particle getting through; and it is also the ratio $(\Psi\Psi^*)_{\text{trans.}}/(\Psi\Psi^*)_{\text{inc.}}$. This statement, however, is too narrow, being valid for the case where the speeds v and v_r of the particles on the two sides of the hill are the same. In the general case, we have:

Probability of transmission or penetration

$$= (v_r/v)(\Psi\Psi^*)_{\text{trans.}}/(\Psi\Psi^*)_{\text{inc.}}$$

$$= (v_r/v)\cdot\phi(E)\cdot\exp.\left[-(2\pi/h)\int_{x_1}^{x_2} dx\sqrt{2m(E-neV_m)}\right]. \quad (27)$$

In Gamow's approximation the product of the first two factors has the pleasantly simple constant value of 4. In the approximations of Gurney and Condon and of Fowler and Nordheim for the cases of Figs. 14 and 15, the product is some function of E which the reader can construct from the foregoing equations. In all these cases, however, it is the exponential factor which dominates the trend of either member of (27) considered as function of E.

Now immediately one sees, that *if* transmutation is due to the penetration of a charged particle through a potential-hill or potential-barrier surrounding a nucleus—and *if* this penetration is governed by laws of quantum mechanics as illustrated in the one-dimensional cases —*then* when the number of observed transmutations is plotted against the kinetic energy K of the impinging particles, the curve should be expected to rise with a gradual smooth upward curvature from the axis of K; and there should be no critical minimum value of K for the advent of the phenomenon, but rather the beginning of perceptible transmutation should be observed at progressively lower and lower energy-values, as the sensitiveness of the detecting-apparatus is improved; and it may well be that transmutation can be detected when K is still so low, that the quotient of K by the charge of the particle is far smaller than any reasonable guess that can be made of the height of the barrier. All these are features of such curves as those of Fig. 17, or Figs. 16 and 17 of the Second Part. The adoption of quantum mechanics permits us to accept these features without ascribing them to the hypothetical functions denoted by f_1 and f_2 in equation

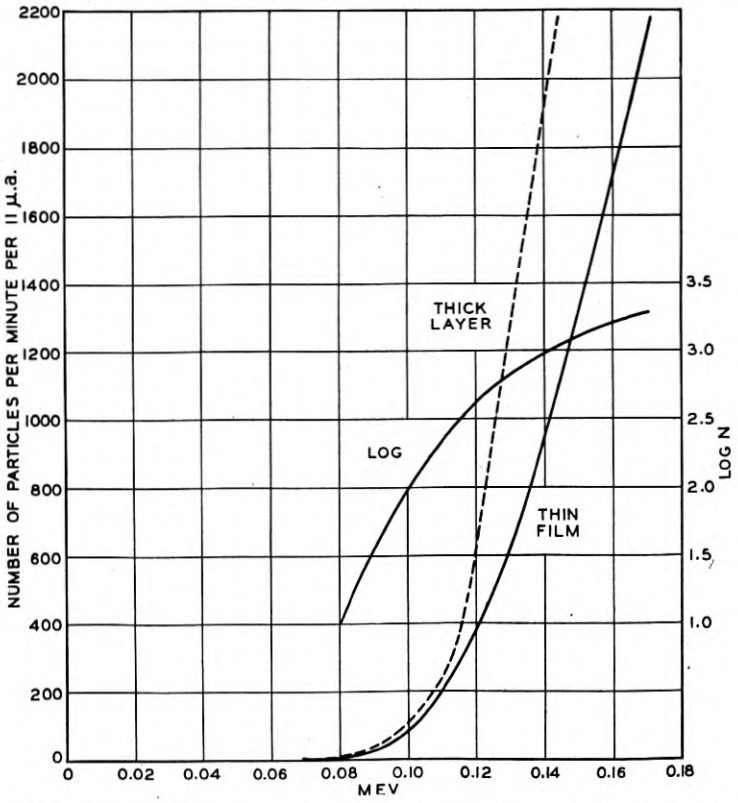


Fig. 17—Transmutation of boron by impact of protons: rate of observed transmutation as function of K, for a very thin film and for a thick layer. (Oliphant & Rutherford, *Proc. Roy. Soc.*).

(20), though it does not rule out the possibility that these functions may have influence upon the curves.

But not all of the curves of probability-of-transmutation versus K are of the simple type of Fig. 17. There are also some which show distinctly-marked peaks superimposed upon the gradual upward sweep: that of Fig. 18 for example, which relates to the transmutation of beryllium by impact of alpha-particles with emission of neutrons, presumably by the process

$$_2He^4 + {}_4Be^9 = {}_6C^{12} + {}_0n^1$$

and that of Fig. 19, which relates to the presumptive process,

$$_2He^4 + {}_{13}Al^{27} = {}_{14}Si^{30} + {}_1H^1.$$

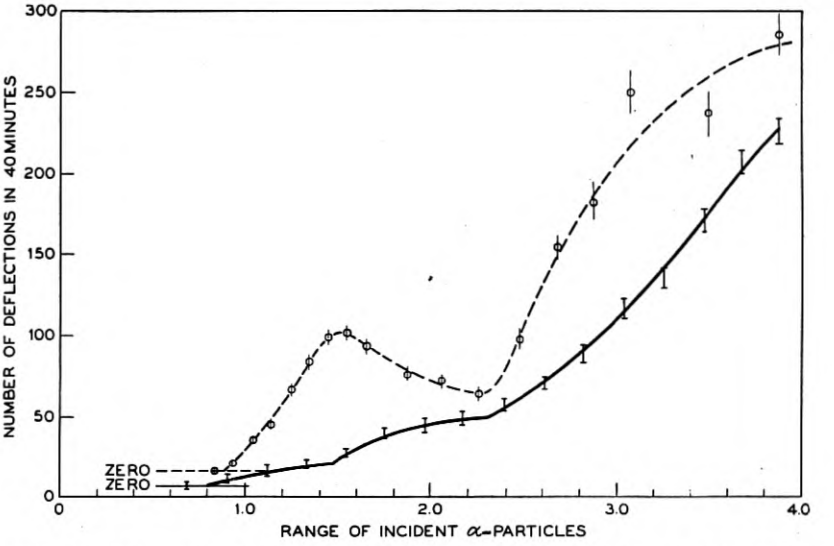


Fig. 18—Transmutation of beryllium by impact of alpha-particles, with production of neutrons; rate of observed transmutation as function of K, for a very thin film and for a thick layer (dashed and full curves respectively), illustrating resonance. (Chadwick; *Proc. Roy. Soc.*).

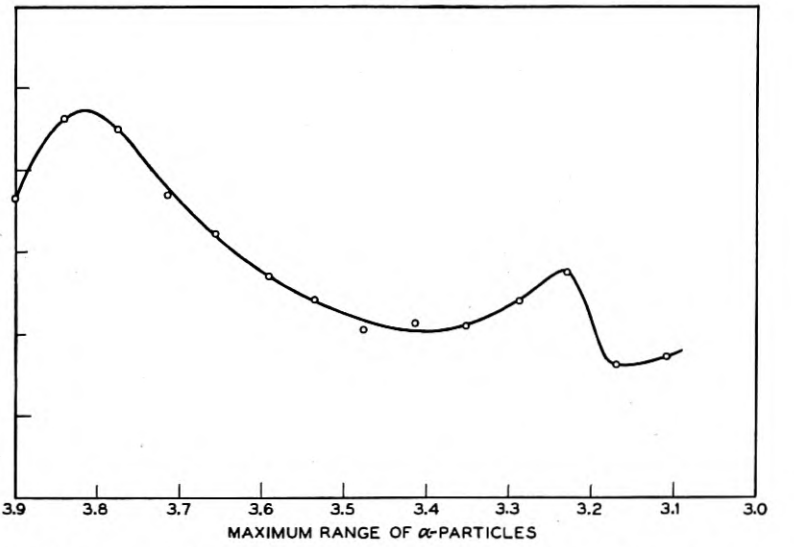


Fig. 19—Transmutation of aluminium by impact of alpha-particles, with production of protons; rate of observed transmutation as function of residual range of alpha-particles, illustrating resonance. (Chadwick & Constable; *Proc. Roy. Soc.*).

In Fig. 19 the abscissa is not K, but a quantity (the range of the impinging alpha-particles) which increases more rapidly than K; but this does not affect the meaning of the peaks. Moreover, there is abundant indication that quantities of such curves are simply waiting for someone to take the data and plot them; for this is the phenomenon of "resonance" to which many pages [27] were devoted in the Second Part, and which has chiefly been observed by the other methods there described, but should always manifest itself in this way when the proper experiments are performed.

If we wish to interpret this without letting go of the classical theory, we must say that either or both of the functions f_1 and f_2 have maxima at certain values of K. But here again, the adoption of quantum mechanics may make this step superfluous. For consider the one-dimensional potential-distribution of Fig. 16, a valley between two hills, with energy-values reckoned from the bottom of the valley. If the wave-equation be solved for this potential-distribution and for any such value of particle-energy E as the dashed line of Fig. 16 indicates—such a value, that according to classical theory a particle possessing it might either be always within the valley or always beyond either hill, but never could pass from one of these three zones to another—a curious result is found. For the solutions which the laws of quantum mechanics demand and accept, the ratio of squared-amplitude $\Psi\Psi^*$ within the valley to squared-amplitude $\Psi\Psi^*$ beyond either hill is usually low, but for certain discrete values of E it attains high maxima!

Now the three-dimensional nucleus-model of which I am speaking resembles this case more than it does the other one-dimensional cases of Figs. 14 and 15, because it consists of a potential-valley surrounded on all sides by a potential-hill. One may therefore expect the probability of entry or penetration to pass through maxima such as are symbolized by the dips in the "fence" of Fig. 13, entailing maxima in the curve of probability-of-transmutation P_t plotted as function of K. Such is the quantum-mechanical explanation of the phenomenon of "resonance," which indeed derives its name from this theory; for the values of K or E at which the maxima occur are those for which the amplitude of the oscillations of the Ψ-function in the valley within the barrier are singularly great.

One wants next to know what quantitative successes have been achieved in predicting or explaining such things as the actual locations of the resonance-maxima, or the precise trend of the curve of P_t-vs-K

[27] "Nucleus, Part II," pp. 148–153; more fully treated in *Rev. Sci. Inst.*, **5**, 66–77 (Feb. 1934).

as it rises away from the axis of abscissæ. Here it must be admitted that almost everything remains to be done. The locations of the resonance-maxima must be expected to depend upon the details of the potential-distribution within the valley, of which there is as yet no notion. The precise trend of P_t as actually observed cannot be the same as that of (27) however good the theory may be, first because not all of the impacts are central (page 596), then because in most (not quite all) of the experiments the bombarded substance is in a thick layer instead of a thin film (page 597), and finally because of the functions f_1 and f_2 of equation (20). Much as we should like for simplicity to put these functions equal to unity and zero respectively, and even though quantum mechanics has removed some of the obstacles to doing so, yet we are obliged to take them into account—the most striking and cogent reason being that with a variety of elements, P_t is not the same for impact of protons as for impact of deutons, though ne is the same for both![28]

The situation being such, one cannot ask as yet for accurate statements about the values of r_m and V_m, the constants of the "crater model" exhibited in Figs. 12 and 13. These must wait upon a thoroughgoing fitting of the theory to the experimental curves of P_t-vs-K, involving a decision as to the magnitude of f_1. The values of V_m for several of the lighter elements have been estimated from the data on transmutation, but the procedure of arriving at the estimates

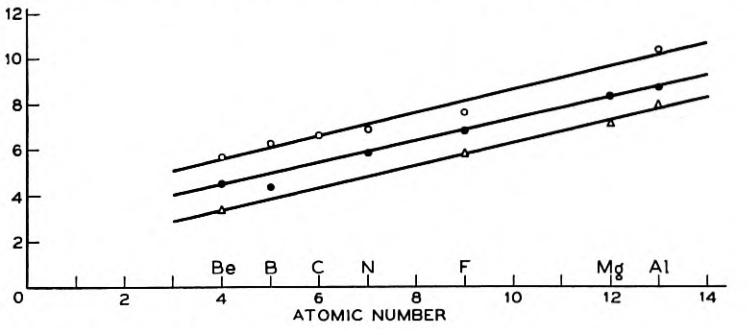


Fig. 20—Resonance-levels and (estimated) heights of potential-barriers for some of the lighter elements, deduced from observations of transmutation. (Pollard; *Phys. Rev.*)

has not (so far as I know) been published. I reproduce as Fig. 20 a graph of Pollard's, the circles along the uppermost line showing the estimated values of eV_m and the crosses along the other two lines

[28] Also it has been said that the shapes of the best experimental curves of P_t-vs-K imply that as K is increased, f_1 increases at first and then becomes constant; but there is a great lack of published theory on these matters.

showing the values of eV at which resonance occurs. The linear trends suggest that these may be properties of the nucleus which are susceptible of simple interpretations. There are also the estimates of V_m made from observations on alpha-particle scattering, most of which are merely minimum-admissible-values below which V_m cannot lie, while a few are more definite. As for r_m, there is at any rate nothing

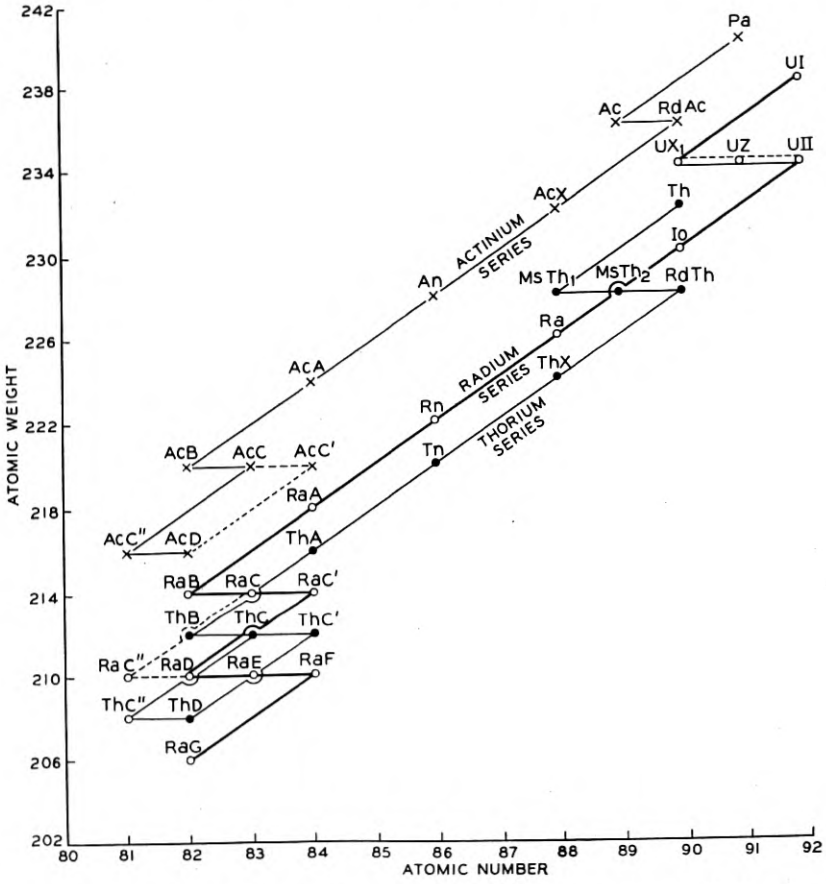


Fig. 21—Genealogies of the radioactive elements. (The actinium series is plotted some distance above the others for legibility, but almost certainly An should lie one unit below Tn, the rest correspondingly).

to indicate that we must make it higher than the values—a few times 10^{-13} cm.—which many reasons impel us to assign to the dimensions of nuclei.

Thus the quantum-mechanical theory of transmutation is as yet

in a primitive state, and indeed not advanced enough (in my estimation) to be considered fully proved by its own successes. Quantum mechanics has, however, many other buttresses, quite sufficiently many to allow us to take it for granted; and in this particular field it has the prestige of prophetic powers. Until the experiments of Cockcroft and Walton, no one had ever effected transmutation except with alpha-particles of charge $2e$ and energy K amounting to several millions of electron-volts. Now Cockcroft and Walton say that they were encouraged to build the elaborate apparatus necessary for trying it with protons of energy much less than one million electron-volts, by Gamow's inference that particles of charge $+e$ should have a very much greater chance of penetrating through a potential-hill and into a nucleus, than particles of equal kinetic energy and only twice the charge—the inference from the fact that ne occurs in the exponent of the exponential function appearing in equation (23) and others like it. Moreover, the phenomenon of resonance was predicted by Gurney (and mentioned by Fowler and Wilson, who, however, apparently did not believe that it could ever be observed) before it was discovered in the experiments of Pose.

The merits of the crater model with the quantum-mechanical theory have, however, not yet been fully presented, for I have left to the last their application to radioactivity.

One of the principal features of radioactivity—both the "induced" variety described in the early part of this article, and the "standard" variety known these thirty-five years—is the exponential decline or decay of the intensity, hence of the quantity of any radioactive substance, as time goes on. This signifies that the average future duration, reckoned from any instant of time, of all the atoms surviving unchanged at that instant, is the same whichever instant be chosen—or, that the probability that an atom, not yet transformed at instant t_0, shall undergo its transformation within (say) a second of time beginning at t_0, has the same value however long the atom may have existed up to this arbitrarily-chosen-moment t_0.

All this is commonly expressed by saying that radioactive transformations obey the laws of chance. I quote (not for the first time) a passage from Poincaré, which illustrates how this had to be interpreted before the advent of quantum mechanics; I take the liberty of writing "nucleus" where he wrote "atom":

". . . If we reflect on the form of the exponential law, we see that it is a statistical law; we recognize the imprint of chance. In this case of radioactivity, the influence of chance is not due to haphazard encounters between atoms or other haphazard external agencies. The

causes of the transmutation, I mean the immediate cause as well as the underlying one (*la cause occasionnelle aussi bien que la cause profonde*) are to be found in the interior of the atom [read, in the nucleus!]; for otherwise, external circumstances would affect the coefficient in the exponent. . . . The chance which governs these transmutations is therefore internal; that is to say, the nucleus of the radioactive substance is a world, and a world subject to chance. But, take heed! to say 'chance' is the same as to say 'large numbers'—a world built of a small number of parts will obey laws which are more or less complicated, but not statistical. Hence the nucleus must be a complicated world." [29]

Well! the advent of quantum mechanics has made unnecessary the conclusion which Poincaré was obliged to draw; for according to this doctrine, the statistical law is characteristic as much of a single particle confronted with a potential-hill, as of the greatest conceivable number of particles mixed up together. It must be admitted that Poincaré's conclusion is probably right enough for the radioactive nuclei of which he knew, all of which must be conceived to comprise several hundreds of particles, protons and electrons and neutrons and the like; but it is not enforced by the reason which he gives, if quantum mechanics is valid. For reversing the argument of previous pages: if in the valley-enclosed-by-hills which is illustrated (for the oversimplified one-dimensional case) by Fig. 16, we postulate a particle and the waves associated with that particle, then the quantum-mechanical boundary-conditions require waves beyond the hills as well, and the coexistence of waves without and within implies a tendency—a tendency governed by the "laws of chance," a probability—for the particle to escape from within to without. As soon as the physicist has successfully made the effort of consenting to quantum mechanics, he is dispensed from the further effort of contriving nuclear models with special features to account for the law of decay of radioactive substances.

Like the probability of entry, the probability of escape of the particle from the confined valley is governed by the ratio of the squared amplitudes $\Psi\Psi^*$ within the valley and beyond the hill (the latter in the numerator). It thus is governed by the exponential function,

$$\exp.\,[-(4\pi/h)\int\sqrt{2m(NeV_m - E)}dx],$$

of which we have already made the acquaintance, multiplied by a

[29] H. Poincaré: "Dernières Pensées," pp. 204–205 (he credits Debierne with the idea).

factor $\phi(E)$ which itself may be a function of E the energy of the particle, but is of secondary importance. Such a formula, in the case of penetration from without, represented the probability of entry for a single approach of the particle to the hill. This suggests that we should deem it in this case as representing the probability of escape for a single approach of the particle from the depths of the valley to the inner side of the hill. Suppose the valley to be of breadth a, the particle to be bumping back and forth in it with speed v_i: the number of approaches of the particle per unit time to the hill will be equal to v_i/a. If the bottom of the valley is at the same level as the axis of abscissæ in Fig. 16, v_i is equal to $\sqrt{2E/m}$; if the valley is deeper, v_i is greater. We deduce for the mean sojourn of the particle within the valley, which is the reciprocal of the probability-of-escape-per-unit-time, the expression:

$$T \geqq [(v_i/a)\phi(E)]^{-1} \exp. [+ (4\pi/h) \int \sqrt{2m(NeV_m - E)}dx]. \quad (28)$$

The aspect of this expression is far from encouraging to one who wishes for a striking quantitative test of the theory. Its value depends not merely on the breadth a assumed for the space within the potential-hill and the height V_m of the hillcrest, but on the details of the shape assumed for the potential-curve of Fig. 12 both within and without the crest; and since there is little or no independent knowledge of these qualities of the nucleus, they may be adjusted practically at will to fit any observed value of T whatever. Furthermore it was obtained by making certain crude assumptions and certain not very close approximations.

One essential test, however, can be applied to it, which it must pass; and pass it does. Let values of a and V_m, and a shape for the potential-hill of Fig. 12, be so chosen that for some particular radioactive element, RaA for instance, equation (28) agrees with experiment; which is to say, that when into the right-hand member of (28) is substituted for E the observed kinetic energy of the emerging alpha-particles, the value of this right-hand member becomes equal to the observed mean life of the element. Now let precisely the same values of a and V_m and the same shape of hill[30] be assumed for some other radioactive element, RaC′ for instance; in the right-hand member of (28), let the observed kinetic energy of the (main group of) alpha-particles for RaC′ be substituted for E; and let the value of T be computed. We

[30] Excepting that the two hills should be expected to slope off towards infinity in the manners of the two functions Z_1/r and Z_2/r, where Z_1e and Z_2e stand for the nuclear charges of the nuclei left behind after the alpha-particle departs, and are often (not always) different for two different radioactive substances.

should not expect a perfect agreement, since the two nuclei are not identical; but we should be disconcerted by a sharp disagreement, since both nuclei belong to elements of which the nuclear charges differ at most by only a few per cent and the nuclear masses by little more. A very great disagreement would in fact be gravely injurious to the theory. Making the test, Gurney and Condon found, however that there is no grave disagreement: the theory survives the test.

A very similar test was applied with greater minuteness by Gamow. For each element he assumed a potential-hill having a vertical rise on the inward side, and on the outward side a curved slope conforming exactly to the function $(Z - 2)/r$, where Z stands for the atomic number of the element before the alpha-particle quits it and consequently $(Z - 2)e$ stands for the charge of the residual nucleus. In other words, he postulated a classical inverse-square electrostatic field ("Coulomb field") from infinity inward to a distance r_0 from the centre of the nucleus, and at r_0 a discontinuous potential-fall. This is a potential-distribution distinguished by a single disposable constant, to wit, r_0; for V_m itself is determined by r_0 and Z.

Gamow proceeded to compute what value must be assigned to r_0 in order to achieve agreement between theory and experiment for each of the twenty-three alpha-emitters. Approximations must be made in carrying through the calculations; those which Gamow employed convert equation (28) into this:

$$\log_e T = -\log_e (h/4mr_0^2) + 8\pi^2 e^2 \sqrt{m}\,(Z - 2)/h\sqrt{2E} + (16\pi e\sqrt{m}/h)\sqrt{Z - 2}\sqrt{r_0}.$$

Putting for E the kinetic energy of the alpha-particles and for T the mean lives of the several elements, he evaluated r_0. Had the values proved very different for the various alpha-emitters, it would have spoken ill for the theory; but all the values were comprised between 9.5 and 6.3 times 10^{-13} cm. The order of magnitude is satisfactory; the differences between the several values are by no means disagreeably great; and there are even signs of a systematic upward trend of the values of r_0 with the atomic numbers of the nuclei. The quantum-mechanical theory and the crater model of the nucleus so pass their crucial test.

The Measurement and Reduction of Microphonic Noise in Vacuum Tubes

By D. B. PENICK

The microphonic response of different types of vacuum tubes to the same mechanical agitation covers a 70 db range of levels. Tubes of the same type, on the average, cover a range of about 30 db. These response levels are too sensitive to minute variations in testing conditions to be measurable with any great precision, but values which are reproducible to within 5 db are obtainable with a laboratory test set comprising a vibrating hammer agitator, a calibrated amplifier, and a thermocouple galvanometer indicator. Sputter noise is made measurable by frequency discrimination methods.

Minimum microphonic disturbance under given service conditions is attained by using the less microphonic types of tubes which are available, by selecting the quieter tubes of a given type for use in positions sensitive to mechanical disturbance, and by protecting the tubes from mechanical and acoustic vibration. Examples of quiet triodes are the Western Electric No. 264B (filament) and No. 262A (indirectly heated cathode). Indirectly heated cathode type tubes are intrinsically less microphonic than filamentary types. Further microphonic improvement in the tubes themselves is made difficult by requirements for favorable electrical characteristics. Well designed cushion sockets can reduce microphonic levels by as much as 30 db, and other methods of cushioning, more expensive and less compact, can extend the reduction even farther. Sputter noise can be eliminated almost entirely in most types of tubes by commonly applied design features and manufacturing methods.

A MAJOR problem which has had to be met by every engineer who has designed a high gain amplifier is that of elimination or reduction of noise. Noise of one kind or another, extraneous to the desired signal, is always present in any amplifier, and sets a lower limit on the smallness of the signal which can be amplified without intolerable interference. In many experimental and commercial amplifiers, the technical and economic obstacles to noise reduction necessitate a compromise between inherent noise level and sufficient volume range for ideal reproduction. The possible sources of this noise are numerous and include power supply, faulty contacts, insulation leaks, pick-up from stray fields, and many other disturbing elements. Among the most persistent types of noise, however, requiring particularly careful design for their elimination or satisfactory reduction, are three which originate in the vacuum tubes themselves, namely, fluctuation noise, microphonic noise, and sputter noise.

Fluctuation noise has been treated at some length by Schottky,[1]

[1] W. Schottky, *Ann. der Phys.*, v. 57, p. 541, 1918; v. 68, p. 157, 1922. *Phys. Rev.*, v. 28, p. 74, 1926.

the discoverer of one of its sources and by several other investigators.[2] It is the fundamental noise arising from the circumstance that the electron current is a stream of discrete particles rather than a continuous flow. For ordinary types of low power tubes of good design, over the audio band of frequencies, the root-mean-square amplitude of this noise is equivalent to about 1 microvolt (120 db below 1 volt) of noise voltage applied to the grid of the tube. G. L. Pearson,[3] in a recent paper, has pointed out that for the best signal-to-noise ratio, the input impedance should be so large that the thermal noise arising in this impedance predominates over the fluctuation noise arising in the tube. In many broad-band or high-frequency systems, however, such an ideal condition is practically unattainable, and fluctuation noise remains as a limiting factor.

It is with the second type, microphonic noise, that we are particularly concerned here, though sputter noise is also of interest and will be dealt with briefly in a later paragraph. Microphonic noise, as the name is usually applied, is the familiar gong-like sound which is always produced when a vacuum tube, followed by a sufficiently high-gain amplifier and a sound reproducer, is subjected to a mechanical shock. Its origin is in the vibrations of the various elements of the tube, which make minute, more or less periodic changes in the spacings of the elements and therefore make corresponding changes in the plate current, whose value at every instant depends on these spacings. Its intensity in a given tube depends on the type and intensity of agitation to which the tube is subjected. For a given agitation, microphonic noise may be reduced either by stiffening and damping the tube structure, thereby reducing the amplitude and duration of vibration of the elements, or by cushioning the tube so that it receives only part of the original agitation.

In order to treat the problem of noise reduction intelligently, it is necessary to have a measure of the effectiveness of treatments applied. To this end, the properties of microphonic response in vacuum tubes have been studied, and a test set has been designed and built for laboratory use which affords a quantitative measure of microphonic response in tubes, and of effectiveness of cushioning in cushion sockets. Some of the more important characteristics of microphonic noise will now be considered.

[2] "The Schottky Effect in Low Frequency Circuits," J. B. Johnson, *Phys. Rev.*, v. 26, pp. 71–85, July, 1925.
"A Study of Noise in Vacuum Tubes and Attached Circuits," F. B. Llewellyn, *Proc. I.R.E.*, v. 18, pp. 243–265, Feb., 1930.
"Shot Effect in Space Charge Limited Currents," E. W. Thatcher and N. H. Williams, *Phys. Rev.*, v. 39, pp. 474–496, Feb. 1, 1932.

[3] "Fluctuation Noise in Vacuum Tubes," G. L. Pearson, *Physics*, v. 5, p. 233, September, 1934. Also published in this issue of *Bell Sys. Tech. Jour.*

Factors Affecting Microphonic Noise Levels

In the first place, it may be pointed out that the production of microphonic noise in commercial types of vacuum tubes is an extremely complicated phenomenon. Each individual component of the mechanical structure is a complete vibrating system having several modes of vibration and natural resonant frequencies, and usually very little damping as compared with electrical circuits. These components, all coupled together mechanically in various ways, form a mechanical network much more complex than the electrical networks encountered in communication engineering practice.

The complexity of the mechanical vibration is reflected in the complex character of the noise itself, and is admirably illustrated by the frequency-response characteristics published by Rockwood and Ferris,[3] and by similar characteristics obtained in the course of this work. It is further demonstrated by the experimental fact that when a large group of supposedly identical tubes is tested by applying the same mechanical vibration to each tube in turn, mounted in the same socket, the response levels of individual tubes may differ from each other by as much as 30 db for representative types of tubes. Such a magnitude of variation would not be expected to result from the comparatively small dimensional variations tolerated in manufacture, and must be explained by the exaggerating effect of intercoupled mechanical resonances in a complicated vibrating system. A curve showing a typical distribution of microphonic response levels in a group of tubes of the same type measured under identical conditions of agitation is shown in Fig. 1. The general shape of this curve is characteristic of any function subject to random variations about a mean, and the range of levels included between the quietest and the noisiest tubes, approximately 30 db, is about average for different types of tubes. Tubes of exceptionally firm construction and of fairly wide spacing, may vary over as small a range as 15 or 20 db, while tubes in which there is a possibility of slight looseness of parts, may vary over as great a range as 40 db or even more.

The nature and intensity of the agitation, the vibrational characteristics of the tube mounting, and the type and degree of mechanical coupling between the tube and its mounting also play an important part in the determination of the microphonic response of a particular tube, since the tube and its closely coupled mounting make up a single vibrating system. The reason for considering the coupling apart from the mounting is that it is usually of the pressure-friction variety and is subject to random variation.

[3] "Microphonic Improvement in Vacuum Tubes," Alan C. Rockwood and Warren R. Ferris, *Proc. I.R.E.*, v. 17, pp. 1621–1632, Sept., 1929.

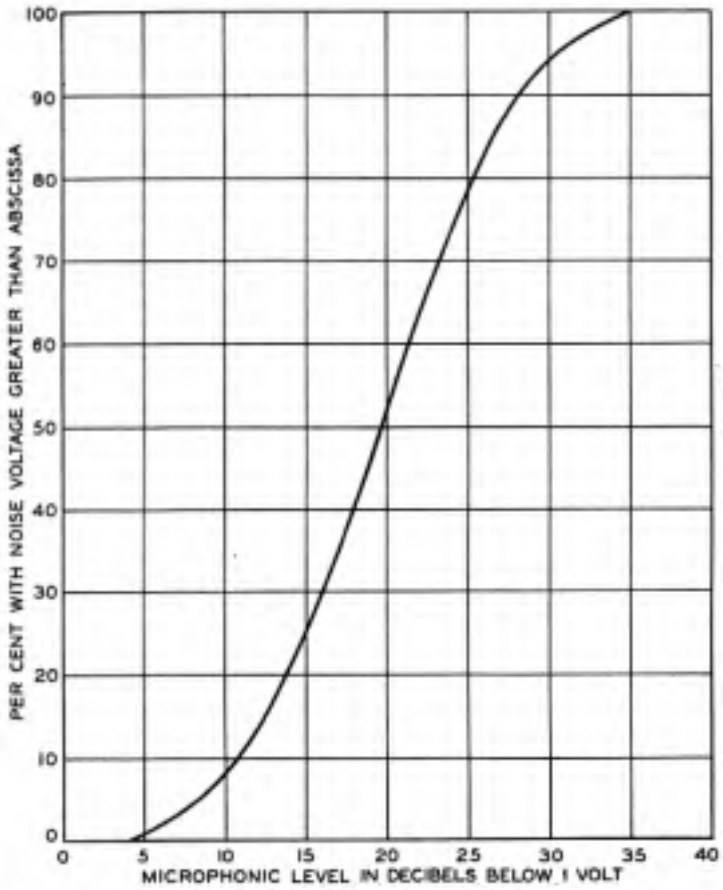


Fig. 1.—Typical distribution of microphonic noise levels produced by a constant, artificial, mechanical stimulus (Western Electric No. 102F Vacuum Tube).

It is found experimentally that with no type of commercial socket which has been tested can a tube be removed and reinserted, or even be left in the socket for a period of time subject to incidental jars and temperature fluctuations, with the expectation of perfectly reproducing a previously measured microphonic level. The sort of random variation which is usually found is illustrated in the reproducibility chart of Fig. 2. In this chart, each point represents two separate observations of microphonic level made on the same tube in the same apparatus, the tube having been removed and reinserted between the two observations. The two levels thus measured are represented by the abscissa and ordinate, respectively, so that if the measurements were perfectly reproducible, all of the points would lie on a straight line making an angle of 45 degrees with the coördinate axes and passing through the origin. The amount of maximum scattering here is about 5 db and may be considered an average value. In some cases, with commercial

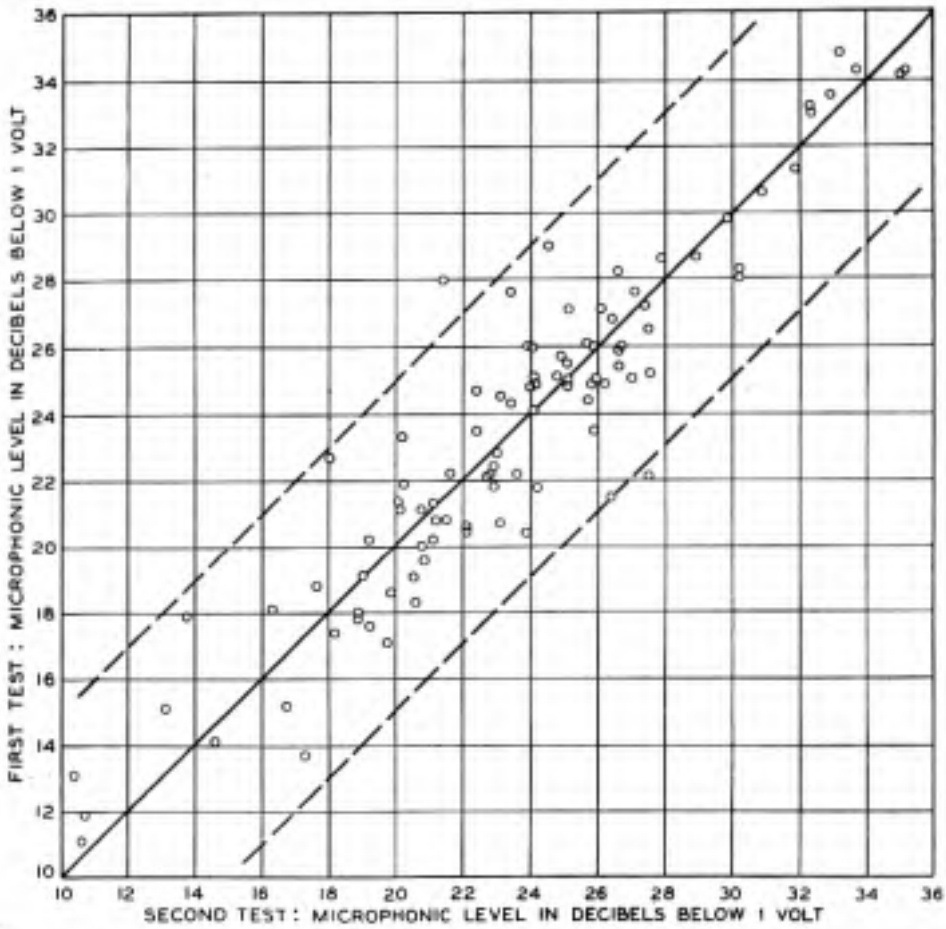


Fig. 2—Reproducibility of microphonic noise level measurements using a commercial socket with a constant, artificial, mechanical stimulus (100 No. 102F Tubes).

sockets, it has been observed to be as low as 3 db and in others as high as 8 db.

In order to show that this random variation is not due to the tube itself, experiments have been made with two forms of suspension which minimize the reaction of the mounting on the vibration of the tube and so reduce as far as possible the effect of variation in coupling. In one set-up, the tube is hung by a single thread of rubber, stretched to its elastic limit, the electrical connections being made by very light, flexible leads fastened with light clips directly to the prongs of the base. In the other, the tube is clamped lightly between two large blocks of very soft sponge rubber, and the electrical connections are made through mercury cups into which the base prongs dip. In both cases, the agitator is a light pendulum striking the base or bulb of the tube. The two mountings give very similar results, and are charac-

terized by very much less scattering than any normal tube mounting, as may be seen in the correlation chart of Fig. 3, which is typical of all of the tests made with these light suspensions. The maximum scattering here is only about 1 db.

Going to the opposite extreme in tube mounting, similar tests have been made with the tube base held tightly in a split metal clamp,

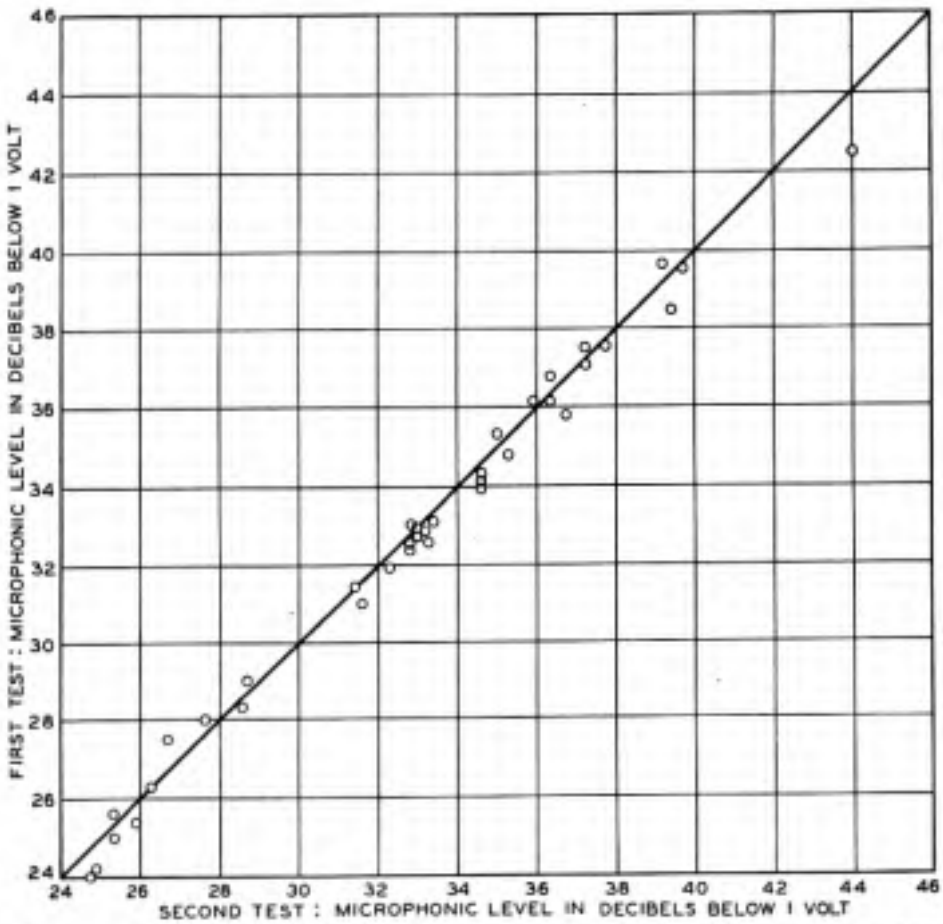


Fig. 3—Reproducibility of microphonic measurements using a rubber clamp tube mounting with a constant, artificial, mechanical stimulus (37 No. 102F Tubes).

which itself is bolted rigidly to a heavy base. As is to be expected, the observed levels vary widely and erratically for successive insertions of the tube, and the mere tightening or loosening of the thumb-screw controlling the pressure of the clamp on the base in some cases changes the level by as much as 10 db.

As for the nature of the applied agitation and the vibrational characteristics of the tube mounting, a countless number of combinations

of these exists, each of which would agitate the tube in a different way. However, from tests made with a variety of mounting arrangements for the tube under test and a variety of degrees of intensity and points of application of forms of impact agitation, it may be concluded that in practical set-ups these factors may be varied widely without changing the general nature of the microphonic level measurements greatly. That is, the form and breadth of the distribution curve and the scattering of the points on the reproducibility chart for any typical group of tubes are likely to be quite similar to Figs. 1 and 2, respectively, for almost any practical impact agitator.

Although the general nature of the results obtained with various combinations of these agitator and mounting arrangements is about the same for all of them, there are certain particular differences, which show up chiefly in two characteristics. One is that the mean noise level of a group of tubes is in general not the same for different mountings and methods of agitation. That this must be true is fairly obvious and needs no comment. The other is illustrated in Fig. 4, which is a correlation chart showing typical results of measurements

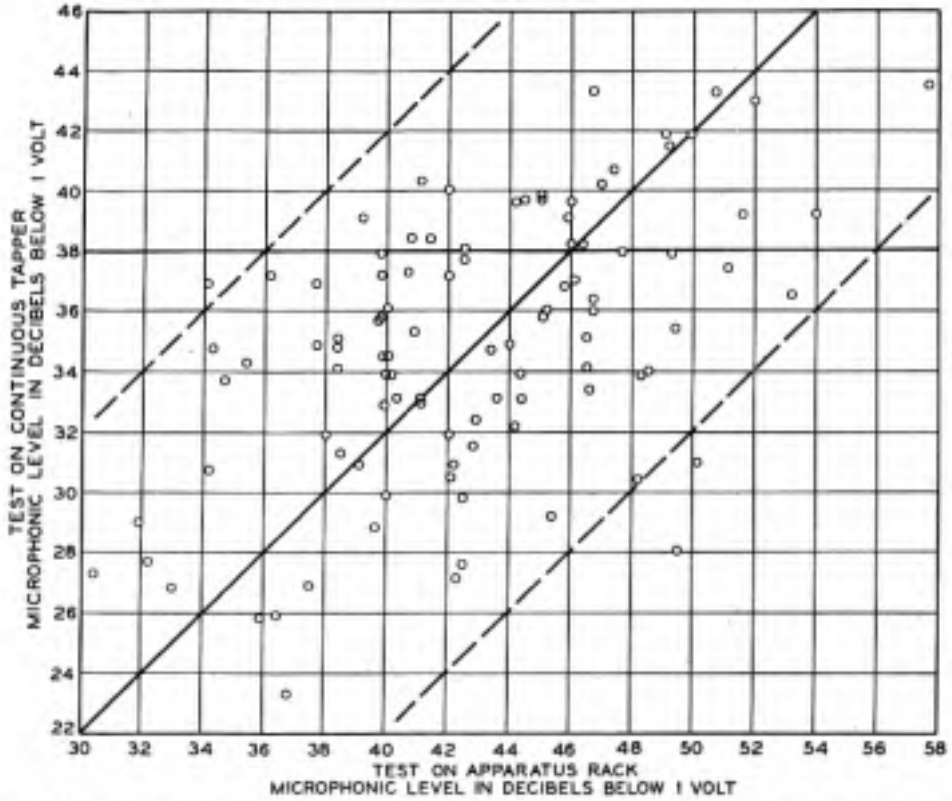


Fig. 4—Comparison of two tube mountings with a constant, artificial, mechanical stimulus (100 No. 102F Tubes).

of the same group of tubes on two different agitating systems. One system in this case consists of a rectangular slate block vibrated by repeated blows of an electrically operated hammer. The other system consists of a steel panel carrying the tube under test, mounted on an apparatus rack which is vibrated by a single blow from a steel ball falling as a pendulum against the rack. The points on this chart scatter about an ideal line over a band about twice as broad as that in Fig. 2 where a test is made and repeated on the same testing unit. It may also be observed that the mean noise levels produced by the two systems are different, about 35 and 43 db below one volt respectively.

The effect of varying the intensity of agitation is shown in Fig. 5.

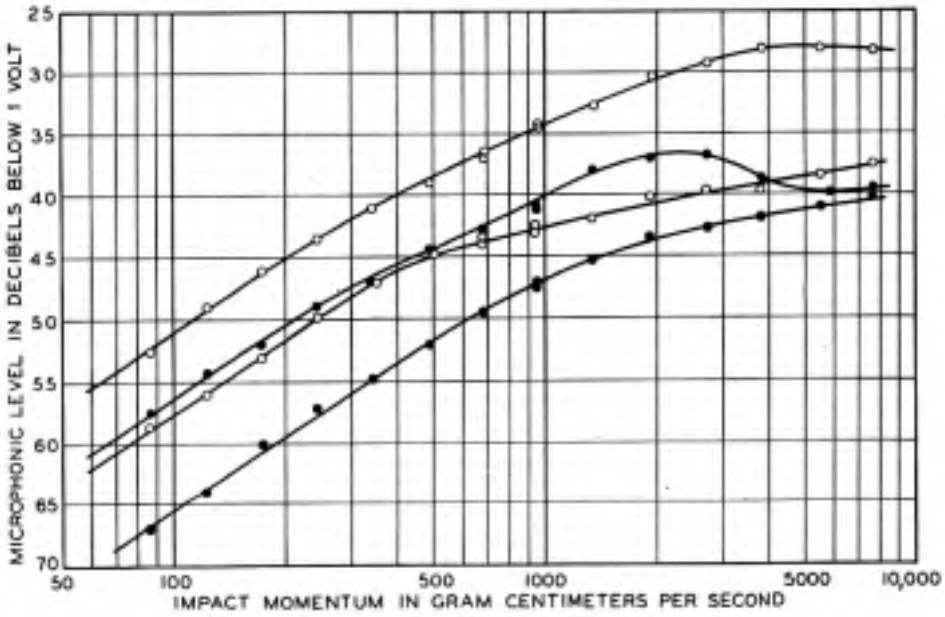


Fig. 5—Effect of intensity of agitation on 4 No. 264B Tubes.

The four curves represent four No. 264B Tubes tested under the same conditions. In making the measurements, the tube under test is mounted in an ordinary socket on a heavy base, which is agitated by means of a pendulum swinging against it and making one rebound. From measurements of the initial swing of the pendulum, its rebound, and its mass, the total momentum imparted to the tube mounting during the impact can be calculated. This quantity is plotted as abscissa in the figure and is proportional to the initial velocity imparted to the tube mounting at the point of impact. Different values of momentum are obtained by varying the initial swing and the mass of the pendulum. At the lower values of momentum, the observed

points lie, within the limits of experimental error, on parallel straight lines so drawn that along them the microphonic noise level expressed in volts is proportional to the initial velocity of the tube mounting. Some such relation as this would be expected to hold as long as the response of the system is linear. The departure from this law at higher values of momentum, then, probably indicates non-elastic motion either of elements of the tube with respect to one another or of the tube with respect to the socket. It may be noted in passing that the No. 264B Tube is exceptionally rigid in structure and that in more loosely constructed tubes, the straight line part of the response curve ends at much lower intensities of agitation.

Since the noise energy is spread over a band of frequencies, the microphonic response observed in any given reproducing system depends also on its frequency-response characteristic. In the usual type of volume indicator, the response is substantially uniform over the audio range of frequencies, but where the final auditory sensation is being considered, the overall characteristic is modified by that of the ear of the listener.[4] The effect of changing the overall response characteristic is illustrated for one particular case in Fig. 6 and Table I.

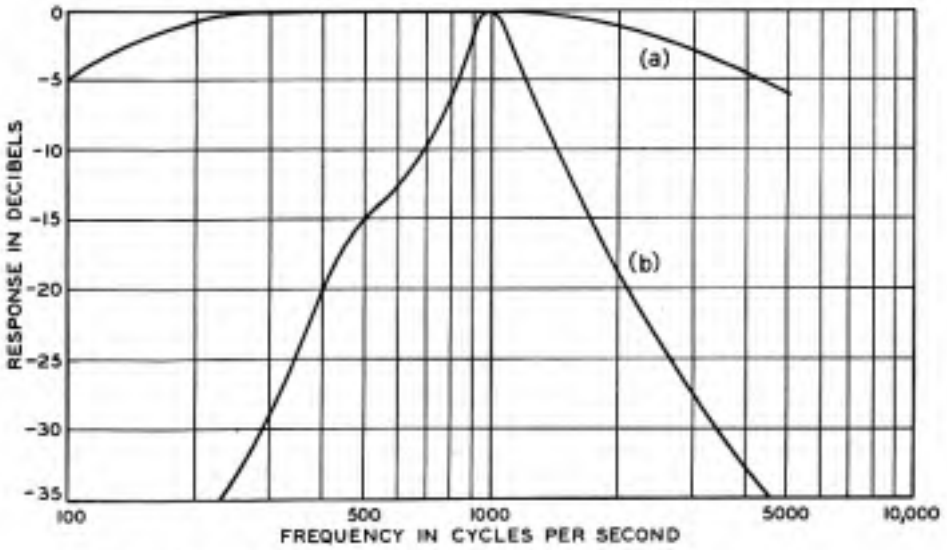


Fig. 6—Microphonic noise amplifier frequency characteristics.

Here, two sets of measurements have been made on each of three types of tubes, one set using an amplifier having a fairly uniform gain characteristic, curve (a), Fig. 6, and the other set using a weighted amplifier, weighted as in curve (b) in the figure. The same agitator and

[4] "Speech and Hearing," H. Fletcher. D. Van Nostrand Co., 1929.

TABLE I

MICROPHONIC LEVELS IN DB BELOW 1 VOLT

Type Tube	Amplifier (a)	Amplifier (b)	Difference
264B	37.9	63.2	25.3
102F	29.7	44.7	15.0
231D	18.2	38.9	20.7

indicator are used in both sets of measurements. Table I gives the mean noise levels for each of the three types of tubes and the differences between the values obtained for each type with the two amplifiers. The results represent about ten tubes of each type. The weighted amplifier, of course, gives the lower levels for all tubes since the noise components at all frequencies except 1000 c.p.s. are amplified less by this amplifier than by the more uniform amplifier. The magnitude of the difference in level depends on the frequency spectrum of the microphonic noise being measured and in general is different for different types of tubes as in this illustration.

Still another important factor which affects the microphonic response of a given system is the relation between the rate of variation of the noise intensity and the time-response characteristic of the system as a whole, usually determined by the indicator. The indicator may be a meter, oscillograph, or other device, or it may be the ear of a listener. A slow moving indicator would respond less to a pulse of noise, such as might be produced by a single shock to a tube, than a more rapidly responding indicator having the same sensitivity to a steady signal. The time required for the ear to reach its maximum response to a suddenly applied sound is about 0.2 second.[5]

The degree of importance of the time-response characteristic of the indicator in measuring transient pulses may be inferred from Table II. This table gives the results of two sets of measurements made on the same three groups of tubes with a single impact type of agitator. The

TABLE II

MICROPHONIC LEVELS IN DB BELOW 1 VOLT

Type Tube	0.2 Second Indicator	2.0 Second Indicator	Difference
264B	37.9	50.5	12.6
102F	29.7	37.2	7.5
231D	18.2	25.2	7.0

[5] "Theory of Hearing; Vibration of Basilar Membrane; Fatigue Effect," G. V. Bekesy, *Physikalische Zeitschrift*, v. 30, p. 115, March, 1929.

two sets differ only in the indicators used. One requires approximately 2 seconds to reach its maximum deflection with a steady impressed signal, and the other requires about 0.2 second. The differences in level corresponding to different types of tubes do not vary greatly, but are nevertheless appreciable. They are, to some extent, a measure of merit of the tube, for a larger difference indicates higher damping of the microphonic disturbance, and high damping is of course desirable.

A Microphonic Noise Measuring Set

The type of test set which has been built in the course of this study for use in the laboratory, comprises an arbitrary standard of agitation, a calibrated amplifier, and an indicating instrument. The agitator consists of a heavy, rectangular slate base at one end of which are mounted sockets for several types of tubes. At the other end is an electrically driven vibrating armature carrying a hammer which strikes about 9 blows per second against a steel block bolted firmly near the center of the slate base. This unit is set on a thick felt pad in a felt lined copper box which provides electrical shielding and some degree of sound-proofing. The sockets used (except those for the bayonet-pin bases) are of the type in which contact springs push each base prong firmly to one side, against the body of the socket. This type has been found to stand up well under repeated insertions and withdrawals of tubes and gives as good correlations between repeated microphonic measurements as any type which has been tried.

The amplifier is basically a simple resistance-choke coupled unit having a frequency-response characteristic which is essentially flat (within 3 db) between 80 and 30,000 c.p.s. The tube under test, whose plate voltage is supplied through an 80-henry choke, works directly into a 100,000-ohm potentiometer, variable in 10 db steps, whose output is connected to the input of the amplifier. The indicator is a sensitive thermocouple galvanometer whose scale is marked off in db and half db divisions so that the noise level may be read directly from the setting of the input potentiometer and the position of the indicator. It has been found convenient to think of the noise level in terms of the root-mean-square voltage developed by the tube across the 100,000-ohm load resistance and to use 1 volt as the reference level. Accordingly, unless otherwise noted, the noise levels given herein are expressed as db below 1 volt across a 100,000-ohm load resistance.

In order to correct for time shifts in tube characteristics and battery voltages, provision is made for checking the amplifier calibration at any time by throwing a switch which transfers its input circuit from the tube under test to a local oscillator. This oscillator delivers a

small, fixed output voltage which is measured and set at a predetermined value with the aid of another thermocouple galvanometer. The amplifier gain may be adjusted, by means of a small range, continuously variable potentiometer until the indicator gives the proper reading to correspond with the known level of the applied input. With reasonably steady battery voltages, this calibration is necessary only two or three times in the course of a day's testing. The range of noise levels for which the amplifier is calibrated extends from 10 db above 1 volt to 65 db below 1 volt. This range has been found to include practically all tubes which it has been desired to test with the standard agitator.

The flat amplifier characteristic, which has been described, is normally used for general testing in connection with vacuum tube design work since it gives the highest microphonic level readings and therefore the most conservative picture of the performance of the tube from the standpoint of the designer. Provision is made, however, for switching in a specially designed weighted amplifier such as is used in making routine noise measurements in telephone speech circuits.[6] The frequency characteristic including this unit has already been shown in Fig. 6, curve (*b*), and is designed to compensate for the interfering effect of each component of noise on the average ear plus the effect of the frequency characteristic of the telephone subset. A similar weighting network compensating for the non-uniform frequency response of the ear alone would also be useful, but has not yet been provided.

Nature and Measurement of Sputter Noise

By making a slight modification of the amplifier circuit, this test set may also be used to measure sputter noise. Sputter noise is a descriptive name applied to a class of noises characterized by a harsh crackling or sputtering sound easily distinguished from the gong-like quality of microphonic noise or the steady roar of electron noise. It may occur either with or without agitation and is the result of discontinuous changes in electrode potential such as may be produced by imperfect contact between conducting members in a tube or by intermittent electrical leaks across insulation.

Sputter noise due to agitation is always accompanied by microphonic noise, and though it often contains instantaneous peaks of high intensity which constitute a very disagreeable and annoying type of interference, its total energy content is usually so small that it contri-

[6] "Methods for Measuring Interfering Noises," Lloyd Espenschied, *Proc. I.R.E.*, v. 19, pp. 1951–54, Nov., 1931.

butes very little to the ordinary microphonic level reading. Special methods must therefore be used in order to make measurements of sputter noise which are independent of microphonic noise. One method which has been found to be effective and convenient, is that of frequency discrimination. If the audio frequency components of the total noise are cut out, then microphonic noise is completely eliminated. Sputter noise, however, due to its discontinuous character has, theoretically, an infinite frequency spectrum, and, practically, one which extends at least into the broadcast band of radio frequencies.

In the microphonic noise test set, sputter noise measurement is provided for by switching in a high-pass filter cutting off sharply at 16,000 cycles. Greater sensitivity is also provided by additional stages of amplification to permit the measurement of the lower levels found to be characteristic of sputter noise in this frequency range. A schematic diagram of the microphonic and sputter noise test set is shown in Fig. 7. The weighted amplifier and the calibrating oscilla-

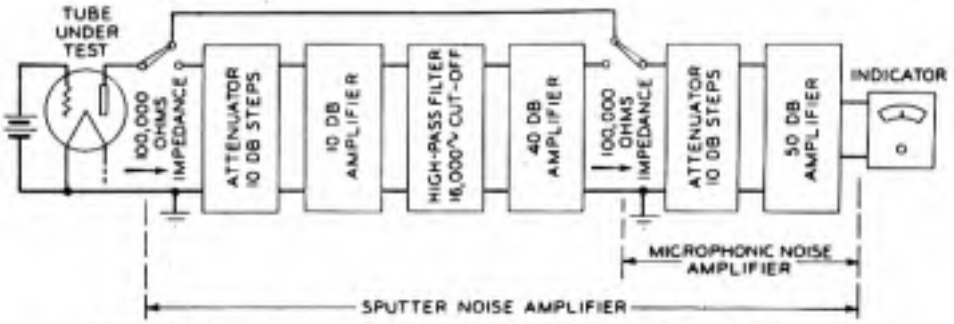


Fig. 7—Microphonic and sputter noise amplifier schematic diagram.

tors (one for microphonic noise and one for sputter noise) have been omitted for the sake of clearness.

Reduction of Microphonic Noise

The reduction of microphonic noise from the view-point of the vacuum tube designer is chiefly a matter of mechanical design and manufacturing technique. The quietest construction is obviously one which has the stiffest electrodes and supporting members, the shortest distances between points of support, and the highest damping of mechanical vibration. The extent to which these features can be incorporated in a practical tube design, however, is limited by the requirements for favorable electrical characteristics. A low filament current, for example, requires that the filament be small in diameter, which renders it more susceptible to vibration than a heavier filament; or if a tube has an indirectly heated cathode, it is a problem to support

the cathode rigidly without conducting away large amounts of heat along the supports. The diameter, length, and spacing of the grid lateral wires are fixed within relatively narrow limits when the desired values of amplification factor and internal impedance are fixed, precluding any important increase in stiffness here; and where high mutual conductance is desired, it is necessary to use relatively close spacings between the elements, under which condition a given amplitude of vibration produces a relatively large per cent change in spacing, and therefore a high microphonic noise level.

The Western Electric No. 264B Vacuum Tube is an example of what has been done in working for a stiff, compact structure. The plate support wires, which also support the whole top of the structure, are short, straight, and as heavy as is practicable, and an extra wire from the press braces the glass bead. One of the most important features of this tube, however, is its filament. In most filament type tubes the vibration of the filament is the chief source of microphonic noise. In the No. 264B Tube, therefore, the filament is made comparatively short and heavy and is mounted in the form of a broad, inverted V to whose apex considerable tension is applied by means of a cantilever spring. The effectiveness of this treatment may be seen from Table III which lists the mean noise levels for a number of types of Western Electric small tubes, and the maximum and minimum

TABLE III*

Class	No. of Samples Tested	Type Number	Microphonic Noise Level in db below 1 volt			
			Measured output Level			Equivalent Mean Input Level
			Max.	Min.	Mean	
Filament type triode	250	101D	23	38	32	47
	833	101F	8	30	19	35
	505	102F	9	36	20	46
	235	215A	12	42	27	41
	1,144	231D	2	28	16	33
	201	239A	4	36	22	37
	715	264B	30	52	42	58
Indirectly heated cathode type triode	99	244A	28	48	39	58
	448	247A	26	52	42	64
	452	262A	36	62	49	71
Screen grid and pentode	24	245A	18	39	29	63
	42	259A	2	36	20	61
	30	283A	4	42	21	62
	30	285A	12	30	23	57

* The microphonic properties of the No. 259A Tube given in this table are identical with those of the 259B discussed by Pearson.[2]

levels which have been observed for each type. The No. 264B, with a mean level of 42 db below 1 volt, is 20 db quieter than the No. 239A which it was designed to replace, and is the quietest of the filament type triodes. The next quietest tube of this structure is the No. 101D, in which the elements are supported from a rigid glass arbor and the filament is quite heavy, requiring one ampere of heating current. The No. 215A is almost identical with the No. 239A except for a firmer supporting structure which results in a 5 db improvement. The most microphonic of the types listed is the No. 231D, which has a very fine wire filament whose diameter is fixed by the requirement that the heating current be 0.060 ampere.

If the filament is the chief source of microphonic noise in filament type tubes, then it is to be expected that tubes having indirectly heated cathodes will be much less microphonic, inasmuch as the cathode is an extremely rigid member. An examination of Table III shows that this is indeed true. The No. 244A and No. 247A types, in which no special precautions have been taken to obtain quietness, are about as quiet as the No. 264B Tube. In the No. 262A Tube, therefore, it has been possible to reduce the microphonic noise still further, to 49 db below 1 volt, by cementing the elements into rigid supporting blocks of ceramic material. This tube is also quiet in other respects, notably in its freedom from AC hum picked up from the cathode heater circuit.[7]

In comparing tubes having widely different electrical characteristics, it is not quite fair to compare their noise output levels alone, for given two tubes having the same noise output, the tube having the higher gain can be used with smaller signal inputs and have no greater noise interference in the output. Accordingly, another column is given in Table III listing the equivalent noise input level which would produce the observed noise output if the tube itself were perfectly quiet. The ratio of this value to the signal input level is directly related to the degree of microphonic noise interference which is effective in the output of the tube. It is computed by adding the voltage gain expressed in db, of the tube in the measuring circuit, to the microphonic output level obtained experimentally. The value of this criterion is illustrated in comparing the noise interference produced by multi-element tubes and triodes. Multi-element tubes as a rule have higher noise output levels than triodes as may be seen by comparing the Nos. 245A, 259A, 283A, and 285A screen-grid and pentode types with the Nos. 244A, 247A, and 262A triodes. When account is taken of the higher voltage amplification of these former types, however, the noise inter-

[7] "Analysis and Reduction of Output Disturbances Resulting from the Alternating Current Operation of the Heaters of Indirectly Heated Cathode Triodes," J. O. McNally, *Proc. I.R.E.*, v. 20, pp. 1263–83, August, 1932.

ference as indicated in the equivalent input noise column of the table, compares quite favorably with that of the triodes.

From the point of view of the user of vacuum tubes, constrained to work with available types, the most effective means of microphonic noise reduction is the use of one of the quieter types of tubes which have been described. In cases where noise difficulties are experienced in existing apparatus not readily convertible to the use of a quieter type of tube, however, some relief may be gained by selecting the quieter tubes from a number of the type to be used. To be fully effective, the selection should be based on measurements made while the tube is in the socket in which it actually works. Under such circumstances, the measurements are reliable to within about 5 db.

Where selection in the field is not feasible, a smaller degree of relief may still be gained by selection at the factory. The degree of effectiveness of this method can be deduced from Fig. 4. Suppose, for example, that quiet tubes for service on the apparatus rack are to be selected by a test made on the continuous tapper. Choosing the best 25 of the group as tested by the continuous tapper (those plotted above the horizontal line in the figure), it is immediately obvious that when these selected tubes are tested on the apparatus rack (compare abscissae in Fig. 4) the worst tubes are somewhat quieter than some of those in the remaining portion of the group. This is more clearly shown in Fig. 8

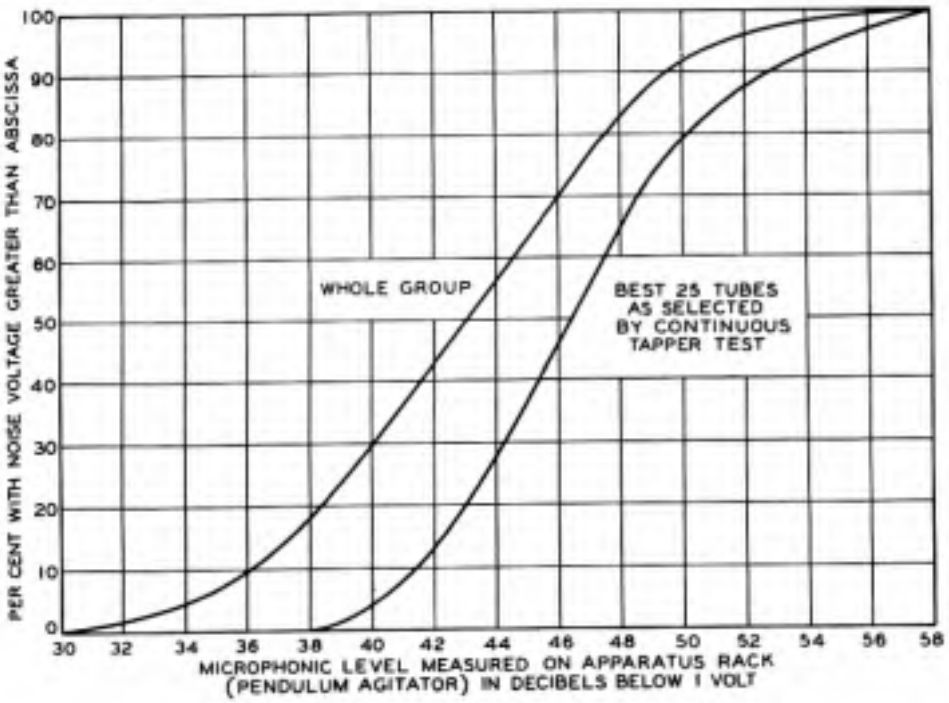


Fig. 8—Effectiveness of selection of quiet tubes.

in which the distribution of levels in this group of 25 tubes is compared with that of the total group as tested on the apparatus rack. The noisiest tubes in the selected group are from 6 to 8 db quieter than the noisiest tubes of the unselected group, and in the selected group, there are none of the tubes which make up the worst 18 per cent of the whole group.

In several commercial situations where microphonic disturbance was at one time troublesome, this type of selection has proved to be of practical value. In these situations, the number of quiet tubes required is only a small percentage of the manufactured output. Furthermore, only a small percentage of the normal output of tubes are found to be prohibitively noisy. Under such circumstances, it is found that when selected tubes from the quietest 25 per cent of the manufacturers stock are used in the positions most sensitive to mechanical shock, the disturbance in these types of equipment either disappears entirely or recedes to such a level that it is no longer troublesome.

Protection from Shock

Where selection of quieter tubes is not feasible or is not sufficiently effective, further reduction of microphonic noise may be achieved by protecting the tube from mechanical and acoustic shock. A very efficient agency for protection from mechanical shock is a well-designed cushion socket. The effectiveness of such a socket depends on its vibrational transmission characteristics considered in relation to the response characteristics of the tubes used. Considerable improvement is usually obtained, however, whatever the combination of tube type and socket type. Figure 9 shows two typical cases of microphonic improvement obtained by using one of several good types of cushion socket which have been tested. The curves drawn in solid lines represent the distributions of microphonic noise levels of a group of No. 102F Tubes tested in one instance in a rigid socket, and in the other in a cushion socket. The mean improvement here due to the cushion socket, is about 30 db. The dotted curves represent similar tests made on a group of No. 262A Tubes and show a mean improvement of about 18 db.

In cases where the noise must be reduced to very low levels, it may not be sufficient to protect the tube from disturbances transmitted mechanically through its base and socket. Except in a perfectly quiet location, there is always some disturbance produced by sound waves impinging directly on the bulb of the tube. Ordinarily this disturbance is negligible, but where the base is sufficiently well cushioned, it may be of controlling importance. It can be reduced only by reducing

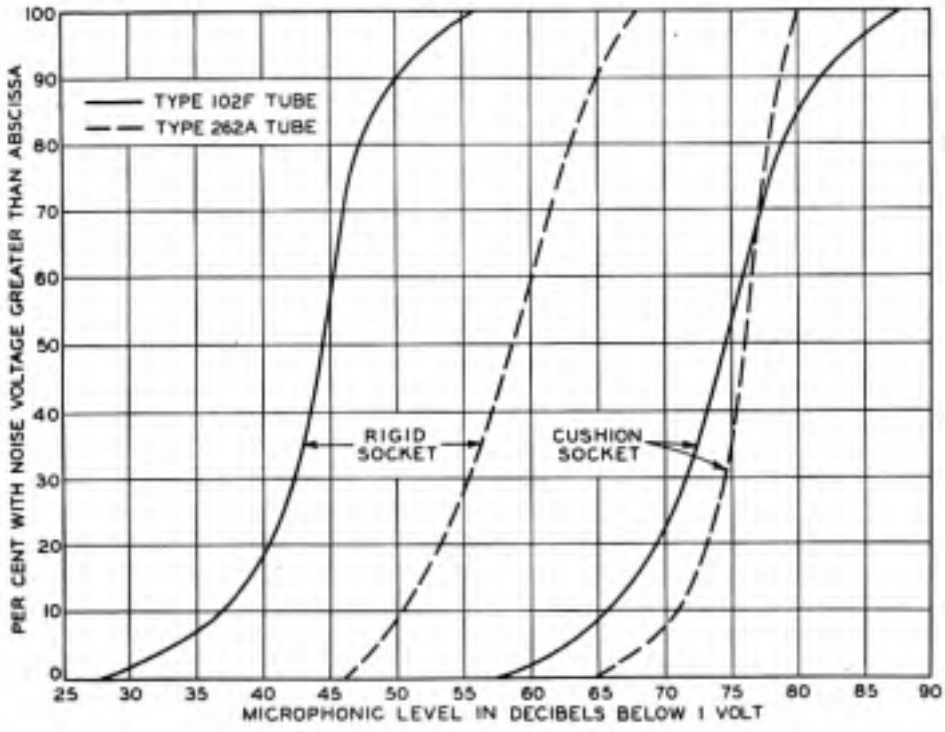


Fig. 9—Effect of cushion socket.

the intensity of the sound wave which is finally allowed to reach the tube, by some such means as enclosing the tube in a heavy, air-tight container.

Reduction of Sputter Noise

The reduction of sputter noise in vacuum tubes is chiefly a problem for the tube manufacturer. Where sputter noise exists in a tube, and exists only with agitation, it is often eliminated by the same cushioning measures which are applied to reduce microphonic noise, but in many cases, satisfactory reduction of sputter would require prohibitive amounts of cushioning. Fortunately, however, the known design features and manufacturing methods, which are now generally applied to tubes of good design, are for the most part quite effective in reducing sputter noise to a negligible level. In the older types of filamentary tubes, for example, sputter noise was often present due to the rattling of the filament at the hook supports at operating temperatures. This source of sputter has been removed in most present day tubes by keeping the filament under tension at all times by means of flexible cantilever spring supports. The effectiveness of this treatment is illustrated in Fig. 10, which shows distributions of sputter noise levels for two

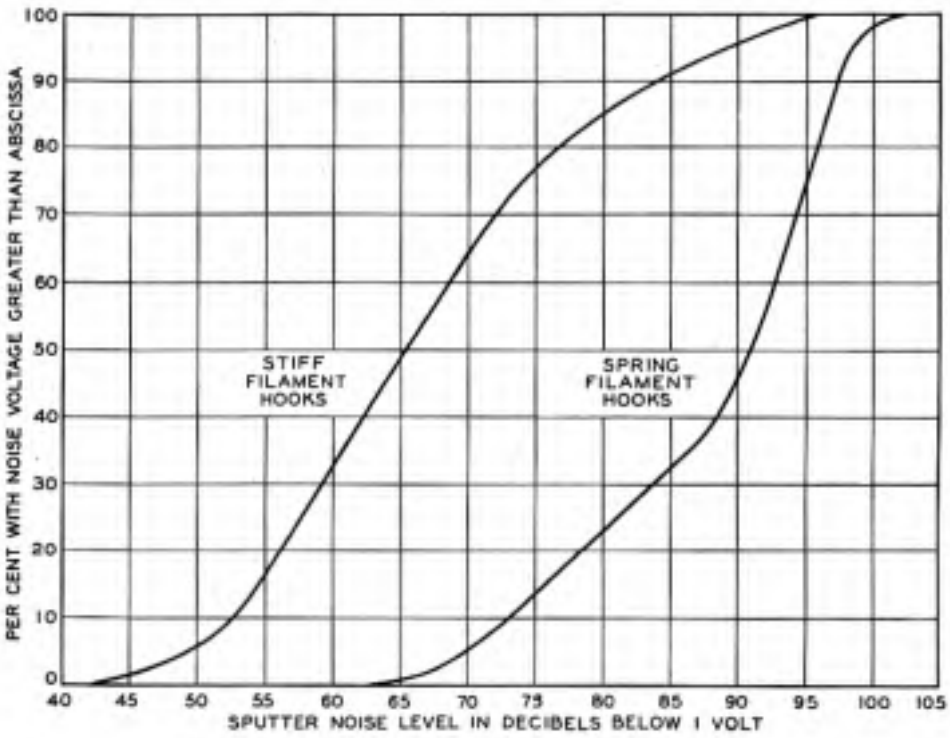


Fig. 10—Effect of filament looseness on sputter noise.

groups of tubes identical in every respect except that one group has spring filament supports while the other has the older rigid supports. In 80 per cent of the tubes, the improvement in the sputter noise is from 20 to 25 db when the spring hook is used.

The source of sputter noise most difficult to control in present day tubes is insulation leaks. These are commonly due to very thin films of conducting material which have been deposited on the surface of the insulating members by sputtering or evaporation during the exhaust or operation of the tube. Experience has shown the conductivity of these films to be intrinsically unstable and discontinuously variable. This condition alone can and does produce sputter noise, but to make matters worse, the metal support wires of the tube are often in only loose contact with the insulating parts and the conducting films covering them so that mechanical agitation breaks and makes the contact and increases the intensity of the noise. The reduction of these insulation leaks is largely a matter of choice of materials, of manufacturing technique to reduce the evaporation of conducting material during exhaust, and of mechanical design to shield important surfaces from contamination during the normal operation of the tube. Great prog-

ress has been made in recent years in effecting an adequate reduction of leaks economically, and in applications where requirements for exceptionally low noise levels warrant slightly increased manufacturing costs, almost any degree of reduction of leaks may be obtained.

Conclusion

The methods which have been outlined for reducing microphonic noise by cushioning and by making use of the quiet tubes which are available are, for the present, adequate to meet all but the most extreme requirements. Should the necessity for further reduction become sufficiently urgent in the future, however, it can probably be obtained either by designing still quieter tubes or by improving the cushioning of sockets. The latter course appears to be the more economical. In either case, however, greatest effectiveness can be attained by considering particular types of tubes and sockets in their relation to one another.

The author is greatly indebted to Drs. M. J. Kelly and H. A. Pidgeon for their kind coöperation and many helpful suggestions in the course of this work.

Fluctuation Noise in Vacuum Tubes *

By G. L. PEARSON

The fluctuation noises originating in vacuum tubes are treated theoretically under the following headings: (1) *thermal agitation* in the internal plate resistance of the tube, (2) *shot effect and flicker effect* from space current in the presence of space charge, (3) *shot effect* from electrons produced by collision ionization and secondary emission, and (4) *space charge fluctuations* due to positive ions. It is shown that thermal agitation in the plate circuit is the most important factor and should fix the noise level in low noise vacuum tubes; shot noise and flicker noise are very small in tubes where complete temperature saturation is approached; shot noise from secondary electrons is negligible under ordinary conditions; and noise from space charge fluctuation due to positive ions is usually responsible for the difference between thermal noise in the plate circuit and total tube noise.

A method is deduced for the *accurate rating of the noise level of tubes* in terms of the input resistance which produces the equivalent thermal noise. *Quantitative noise measurements* by this method are reported on four different types of vacuum tubes which are suitable for use in the initial stage of high gain amplifiers. Under proper operating conditions the noise of these tubes approaches that of thermal agitation in their plate circuits at the higher frequencies and is 0.54 to 2.18 $\times 10^{-16}$ mean square volts per cycle band width in the frequency range from 200 to 15,000 cycles per second. Below 200 cycles per second the noise is somewhat larger.

The minimum noise in different types of vacuum tube circuits is discussed. These include input circuits for high gain amplifiers, ionization chamber and linear amplifier for detecting corpuscular or electromagnetic radiation, and photoelectric cell and linear amplifier for measuring light signals.

With the aid of these results it is possible to design circuits having the maximum signal-to-noise ratio obtainable with the best vacuum tubes now available.

Introduction

IT is well known that the noise inherent in the first stage of a high gain amplifier is a barrier to the amplification of indefinitely small signals. Even when fluctuations in battery voltages, induction, microphonic effects, poor insulation, and other obvious causes are entirely eliminated, there are two sources of noise which remain, namely, thermal agitation of electricity in the circuits and voltage fluctuations arising from conditions within the vacuum tubes of the amplifier. The effect of thermal agitation in circuits outside the vacuum tube is well understood, but in the case of tube noise there is considerable confusion. In order to clarify the whole subject, the present paper analyzes the various sources of noise in vacuum tubes and their attached circuits, points out a new method for the measurement of tube noise, reports the results of such measurements on four different types of vacuum tubes, and discusses the minimum noise in different types of vacuum tube circuits.

* Published in *Physics*, September, 1934.

Often, in the use of high-gain amplifiers, the impedance of the input circuit is naturally high or may effectively be made high by the use of a transformer. In this case the contribution of noise from the vacuum tube is small compared with the noise arising from thermal agitation in the input circuit. This is a desirable condition since it furnishes the largest ratio of signal to noise for a given input power. Sometimes, however, the input impedance is perforce so small that the tube noise may be comparable with or greater than the thermal agitation noise. Such conditions may arise, for example, in amplifiers where the frequency dealt with is high or the frequency range is wide. It is, therefore, desirable to know the noise level to be expected from different types of tubes that may be used in the first stage of high-gain amplifiers as well as to be able to calculate the thermal noise level of the input circuit.

The noise of thermal agitation [1] arises from the fact that the electric charge in a metallic conductor shares the thermal agitation of the molecules of the substance so that minute variations of potential difference are produced between the terminals of the conductor. The mean square potential fluctuation is proportional to the absolute temperature and to the resistive component of the impedance of the conductor, but is independent of the material. The thermal noise power is distributed equally over all frequencies although the apparent magnitude depends on the electrical characteristics of the measuring system as well as on those of the conductor itself. From purely theoretical considerations the following equation has been derived [2] to give the thermal noise voltage at the output of an amplifier due to the thermal agitation of electric charge in an impedance at the input:

$$\overline{E_T^2} = 4kT \int_F R(f) |G_1(f)|^2 df. \tag{1}$$

$\overline{E_T^2}$ is here the mean square thermal noise voltage across the measuring device, k is Boltzmann's constant (1.37×10^{-23} watt second per degree), T the temperature of the impedance expressed in degrees Kelvin, $R(f)$ the resistive component of the impedance at the frequency f, $G_1(f)$ the voltage amplification between the input impedance and the measuring device at the frequency f, and F the frequency band within which the amplification is appreciable.

While the thermal noise in the circuit is accurately predictable, the noise originating within the vacuum tube is not completely under-

[1] J. B. Johnson, *Phys. Rev.*, **32**, 97 (1928).

[2] H. Nyquist, *Phys. Rev.*, **32**, 110 (1928).

stood and cannot be calculated accurately. It is known, however, that tube noise arises from a number of different causes, chief among which are: (1) thermal agitation in the internal plate resistance of the tube, (2) shot effect and flicker effect from space current in the presence of space charge, (3) shot effect from electrons produced by collision ionization and secondary emission, and (4) space charge fluctuations due to positive ions. Each of these sources of noise will be discussed in the following section:

Origin of Noise in Thermionic Amplifier Tubes

Thermal Agitation in the Internal Plate Resistance of the Tube [3]

Just as voltage fluctuations are produced by thermal agitation in resistances comprising the input circuit, so the resistance component of the impedance between plate and cathode is a source of thermal noise. This impedance consists of the internal plate impedance of the vacuum tube in parallel with the external load impedance. Llewellyn [4] has shown that the resistive component of the internal plate impedance produces thermal noise as if it were at the temperature of the cathode. The following formula has been developed by him to cover the case where the tube impedance and load impedance are pure resistances:

$$\overline{E_T^2} = 4k[(r_p r_0)/(r_p + r_0)^2](T_0 r_p + T_f r_0) \int_F |G_2(f)|^2 df. \qquad (2)$$

Here r_p is the internal plate resistance of the tube, r_0 the external load resistance in the plate circuit, $G_2(f)$ the voltage amplification between the load resistance r_0 and the measuring device, and T_0 and T_f respectively the temperatures of the external load resistance and cathode expressed in degrees Kelvin. The relationship between $G_1(f)$ and $G_2(f)$ is given by

$$G_1(f) = G_2(f)(\mu r_0)/(r_0 + r_p), \qquad (3)$$

where μ is the voltage amplification factor of the tube. By assuming $G_2(f)$, in equation (2), to be constant over the frequency range F and substituting for it the value given by equation (3), it is found on integrating that the thermal noise in the plate circuit of the tube produces the same effect in the measuring device as a signal applied

[3] During the preparation of this paper a paper by E. B. Moullin and H. D. Ellis entitled "Spontaneous Background Noise in Amplifiers Due to Thermal Agitation and Shot Effects" appeared in *I. E. E. Jour.*, **74**, 323 (1934). The authors there contend that no thermal noise is produced in the plate impedance of a thermionic vacuum tube and that shot noise is not altered by the presence of space charge. With these contentions I cannot agree and I hope to state my definite reasons therefor at a later date.

[4] F. B. Llewellyn, *Proc. I. R. E.*, **18**, 243 (1930).

to the input circuit whose magnitude at the grid expressed in mean square volts is given by

$$\overline{V^2} = 4kT_0(r_p/\mu)^2[T_f/(T_0 r_p) + 1/r_0]F. \quad (4)$$

Since the noise of thermal agitation is always present, this equation gives the absolute minimum to which fluctuation noise in an amplifying tube can be reduced after all other causes have been eliminated. It shows that for the ideal low noise tube in which thermal noise in the plate circuit is the limiting factor, the noise level may be reduced by a decrease in the cathode temperature, a decrease in the effective frequency band, or by an independent decrease in the plate resistance or increase in the amplification factor. In order to operate at a minimum noise level the tube should work into a load resistance which is large in comparison with $r_p T_0/T_F$. Under this circuit condition the noise level is inversely proportional to μ^2/r_p, a quantity often defined as the "figure of merit" of an amplifying tube.

Shot Effect and Flicker Effect in the Presence of Space Charge

The theory of the shot effect in the absence of space charge has been studied quite completely both theoretically and experimentally by many investigators.[5] The results, however, are not applicable to the study of noise in thermionic vacuum tubes used in high-gain amplifiers, since a high degree of space charge is required in tubes used for this purpose. Llewellyn has extended the theory of the shot effect to cases where partial temperature saturation exists, and obtained a general equation to cover all conditions.[4] This equation reduces to the following form when the load impedance is a pure resistance:

$$\overline{E_g^2} = 2ej(\partial i/\partial j)^2[r_p r_0/(r_p + r_0)]^2 \int_F |G_2(f)|^2 df. \quad (5)$$

$\overline{E_g^2}$ is here the mean square shot voltage across the measuring device, i the total space current, j the total current emitted by the cathode, and e the electronic charge (1.59×10^{-19} coulomb).

A precise experimental verification of this equation is very difficult because of the difficulty in determining $\partial i/\partial j$ accurately. Thatcher,[6] however, has made shot measurements in the presence of space charge ($1 \gtrless \partial i/\partial j \gtrless 0.66$) which verify the theory within the experimental error of the determination of $\partial i/\partial j$.

[5] W. Schottky, *Ann. d. Physik*, **57**, 541 (1918); T. C. Fry, *Jour. Franklin Inst.*, **199**, 203 (1925); A. W. Hull and N. H. Williams, *Phys. Rev.*, **25**, 147 (1925).

[6] Everett W. Thatcher, *Phys. Rev.*, **40**, 114 (1932).

Equation (5) shows that as long as space charge is too small to affect the flow of current, that is when i is equal to j, the mean square shot voltage is directly proportional to the space current. As emission is increased, however, space charge begins to control and finally limits the space current so that the value of $\partial i/\partial j$ approaches zero. Thus the shot voltage increases less rapidly as space charge becomes effective and then finally decreases rapidly toward zero as complete space charge control is reached.

Experimental curves showing the effect of space charge on tube noise are shown in Fig. 1 where abscissæ represent space current in

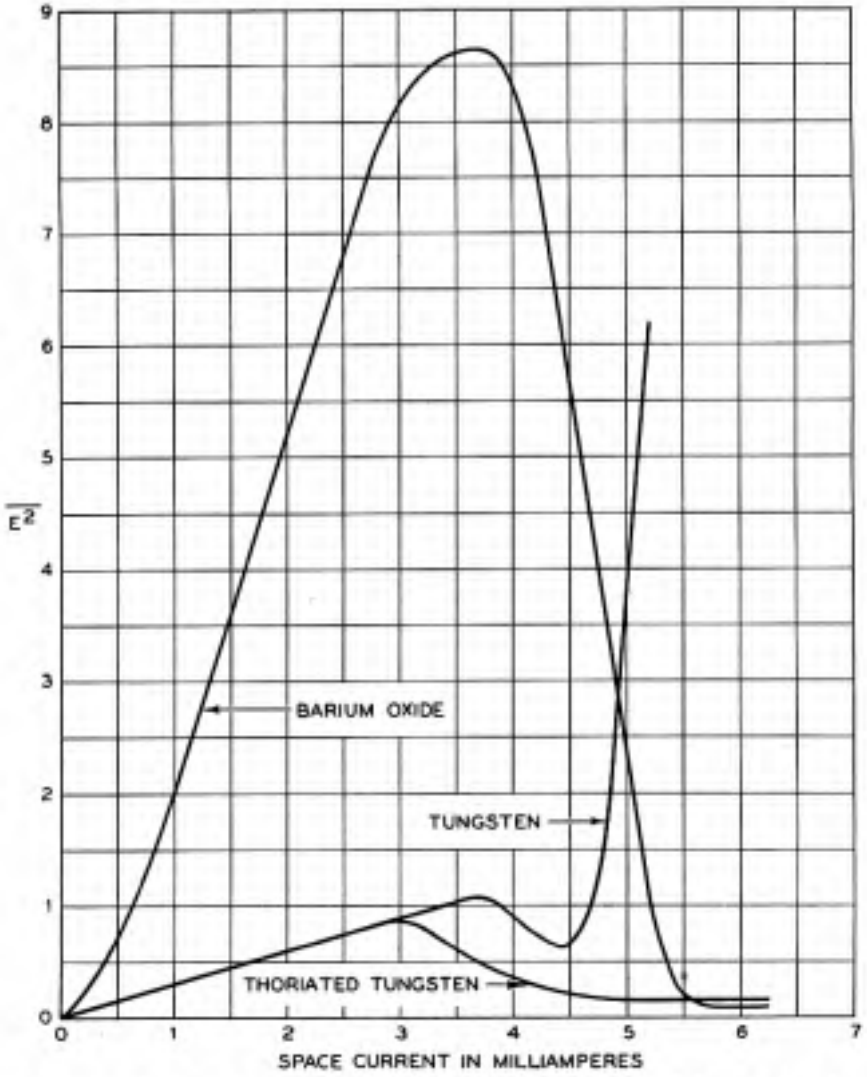


Fig. 1—The effect of space charge on fluctuation noise. Three tubes having filaments composed of tungsten, thoriated tungsten, and barium oxide. $\overline{E^2}$ is the mean square noise voltage across the output measuring device expressed in arbitrary units. The variation in space current was obtained by changing the cathode temperature, the plate voltage remaining constant.

milliamperes, and ordinates represent mean square noise voltage across the output measuring device expressed in arbitrary units. The change in space current was obtained by varying the filament heating current while the plate voltage remained constant. Tubes having thoriated tungsten, tungsten, and barium oxide cathodes were used.

At low space currents where no space charge is present the thoriated tungsten and tungsten filaments each give a pure shot effect, the mean square voltage increasing linearly with the space current. As the space current is increased further and space charge sets in, the shot voltage in each tube goes through a maximum and decreases with oncoming temperature saturation as suggested by equation (5). With the approach of complete temperature saturation the noise, however, does not decrease to zero in accordance with this equation. If it were possible to reach complete temperature saturation the residual noise would not be due to the shot effect, but rather to thermal noise in the plate circuit of the tube, positive ions and secondary emission within the tube, and other contributing causes. Usually this condition is approached in the better commercial tubes so that the contribution of true shot noise is a small part of the total noise.

If the methods used in obtaining equation (4) are applied to equation (5), it is found that the shot noise in the plate circuit of the tube produces the same effect in the output measuring device as a signal applied to the input circuit whose magnitude at the grid expressed in mean square volts is

$$\overline{V^2} = 2ej(\partial i/\partial j)^2(r_p/\mu)^2F. \tag{6}$$

This equation shows that the level of shot noise at the input is lowered by an increase in the cathode temperature, which increases the degree of temperature saturation, and by an increase in the ratio μ/r_p, which by definition is the transconductance of the tube, but is independent of the external load resistance. It should be remembered, however, that shot noise in the plate circuit should not fix the noise level in low noise vacuum tubes and that never, as is sometimes done in the literature, can the noise of an amplifier be calculated as pure shot noise in the plate circuit, for in the absence of space charge the tube would not be an amplifier.

Although space charge can counteract the effect of random electron emission from the cathode so that shot noise is reduced, other factors can alter the flow of current in such a way that the noise is increased. This is the case when changes in emission occur over small areas of the cathode, giving rise to an additional fluctuation which has been termed flicker effect.[7] This type of noise is particularly noticeable

[7] J. B. Johnson, *Phys. Rev.*, **26**, 71 (1925); W. Schottky, *Phys. Rev.*, **28**, 74 (1926).

with oxide coated cathodes. Since the flicker effect is due to localized variations in the emission of the cathode, one would expect it to disappear in the presence of a complete space charge condition.

The experimental curve for the barium oxide coated filament, Fig. 1, shows a flicker effect many times larger than the shot effect on which it is superimposed. At low space currents the mean square flicker effect voltage increases faster than the pure shot noise, a square law rather than a linear relationship being followed. As space charge sets in, the flicker effect voltage goes through a maximum and then decreases with increased space current in the same manner as does the shot effect voltage. In spite of the large flicker effect, as complete temperature saturation is approached the total noise is even less than that found with the thoriated tungsten filament which has no flicker effect. This illustrates clearly the effectiveness of space charge in smoothing the space current.

When the control grid of a vacuum tube is floating at its equilibrium potential, the noise level is much higher than when the grid is connected through an input circuit to the cathode. This increase in noise is primarily due to thermal noise in the extremely high input resistance of the tube and to shot noise arising from small grid currents.[8] The magnitude of the thermal noise may be calculated, knowing that the input impedance of the tube consists of its input resistance, r_g, in parallel with its dynamic grid-to-ground capacitance. In such a combination the real resistance component, $R(f)$, is related to the pure resistance, r_g, and the dynamic capacitance, c, according to the equation

$$R(f) = r_g/(1 + 4\pi^2 c^2 r_g^2 f^2). \tag{7}$$

According to equation (1) the mean square thermal noise input voltage is then

$$\overline{V_T^2} = 4kTr_g \int_F df/(1 + 4\pi^2 c^2 r_g^2 f^2). \tag{8}$$

With the grid floating at its equilibrium position (usually slightly negative with respect to the cathode) the grid current is composed of two components equal in magnitude but opposite in sign. The one component consists of electrons reaching the grid, while the other consists of positive ions reaching and electrons leaving the grid. The electrons are liberated from the grid by secondary emission, the photoelectric effect, thermionic emission, and soft X-rays. It should be pointed out that space charge does not reduce the noise produced by

[8] L. R. Hafstad, *Phys. Rev.*, **44**, 201 (1933).

the shot effect in any of these currents. The general shot effect equations [9] show that the magnitude of shot noise from these grid currents is

$$\overline{V_g^2} = 2ei_g r_g^2 \int_f df/(1 + 4\pi^2 c^2 r_g^2 f^2)^2, \tag{9}$$

where i_g is the sum of the grid currents regardless of sign.

Noise Produced by Secondary Effects

In this classification are grouped several sources of disturbance whose individual effects are very difficult to calculate and measure under the operating conditions of the vacuum tube. For this reason the following discussion will include only a general consideration of the more obvious contributing causes.

Although the cathode is the principal source of electrons which reach the plate, in actual practice electrons are produced by ionization of the gas molecules within the tube or by secondary emission resulting from bombardment of the tube elements. Electrons produced in this manner are drawn to the plate and generate noise which is not much affected by the space charge. Assuming a reasonable magnitude for the current produced in this manner it can be shown by the shot equations that noise from this source is usually negligible. In cases where the gas pressure within a tube is above normal, or in screen-grid and multi-grid tubes having high plate resistances and considerable secondary emission, the shot noise from secondary and ionization electrons may be of the same order of magnitude as thermal noise in the plate circuit.

Positive ions formed from ionized gas molecules or emitted from the tube elements are much more effective in producing noise since, instead of being drawn off to the plate, they are attracted into the space charge region where small disturbances in equilibrium produce large momentary fluctuations in space current. Due to their large mass the motions of the ions are relatively slow, so that they are very effective in this respect. This type of noise is quite disturbing in amplifying tubes for it tends to become a maximum at complete temperature saturation. This is illustrated very clearly in the noise measurements on the tungsten filament shown in Fig. 1. Here positive ions from the filament begin to show their effect as space charge sets in, the number of ions and the amount of noise increasing as temperature saturation is approached. As heard in the loud speaker, this noise consists of sharp crackling sounds which can easily be dis-

[9] E.g. Ref. 4 or 5.

tinguished from the steady rustling noise of the shot and thermal effects.

Ballantine [10] has recently made calculations and measurements on the noise due to positive ions from collision ionization in which he has shown that the mean square noise voltage is roughly proportional to the gas pressure within the tube and to the 3/2 power of the plate current. Comparing his results with equation (2), it appears that under ordinary working conditions the noise due to collision ionization in a vacuum tube may be of the same order of magnitude as noise from thermal agitation in its plate circuit. The noise level of tubes having a poor vacuum, however, may be much higher.

Measurement of Tube Noise

The performance, as regards freedom from noise, of a vacuum tube used in an amplifier may be indicated by a comparison between the noise and a signal applied to the grid. Usually we say that the noise is equivalent to a signal which gives the same power dissipation in the output measuring instrument as the noise, the frequency of the signal being suitably chosen with respect to the frequency characteristics of the amplifier. Since tube noise is distributed over all frequencies and the noise power increases with the effective band width, it will be advantageous to express this input signal in equivalent mean square volts per unit frequency band width, effective over a given frequency range.

From these considerations it can be seen that the most convenient standard signal for measuring the equivalent input noise over any given frequency range is one in which the mean square signal voltage is distributed equally over all frequencies. With such a signal the equivalent input noise over any frequency range can be measured directly, while if an oscillator is used a number of measurements are required and the result must be computed by graphical integration. A signal which meets these frequency requirements perfectly is the noise of thermal agitation. Accordingly, in the measurements to be described here the standard input signal will be the thermal agitation voltage of a resistance R, connected between the control grid and cathode of the tube under test.[1]

The thermal noise voltage of the grid circuit, referred to the output measuring device, is given by equation (1), where $R(f)$ is the real resistance component of an input impedance consisting of the pure resistance R in parallel with its shunt capacity and that of its leads

[10] Stuart Ballantine, *Physics*, **4**, 294 (1933).

and of the vacuum tube. In such a combination $R(f)$ is related to the pure resistance R and the total capacitance c according to equation (7). In all the measurements described here the factor $4\pi^2c^2R^2f^2$ is so small in comparison with unity that it may be neglected without appreciable error. Under these conditions equation (1) reduces to

$$\overline{E_T^2} = 4kTR \int_F |G_1(f)|^2 df, \tag{10}$$

where R is the direct current value of the resistance between control grid and cathode of the tube under test.

The voltage fluctuations arising from conditions within the tube produce a mean square voltage output $\overline{E_N^2}$ according to the equation

$$\overline{E_N^2} = \int_F |V(f)|^2 |G_1(f)|^2 df, \tag{11}$$

where $|V(f)|^2$ is the tube noise at the frequency f for unit frequency band width, expressed in volts squared and referred to the input circuit. Letting V_F^2 be the effective value of $|V(f)|^2$ over the band width of the amplifier we obtain

$$\overline{E_N^2} = V_F^2 \int_F |G_1(f)|^2 df. \tag{12}$$

Since the integrals in equations (10) and (12) are identical it is found on dividing one equation by the other and solving for V_F^2 that:

$$V_F^2 = 4kTR(\overline{E_N^2}/\overline{E_T^2}). \tag{13}$$

Equation (13) enables one to calculate the magnitude of tube noise in the frequency range F, per unit cycle band width, in terms of the thermal noise generated in a resistance R placed in the input circuit.[11] Since this equation contains no integral the measurements are simplified in that neither standard signal generator nor calibrated amplifier is required.

Apparatus

The experimental arrangement used in the measurements to be reported here is given in schematic form in Fig. 2. The system includes the tube under test, a high gain amplifier, appropriate filters, an attenuator, and an output measuring device.

[11] It is assumed that tube noise does not vary with frequency, or that the band width of the amplifier is so narrow that no appreciable error is introduced in applying the result.

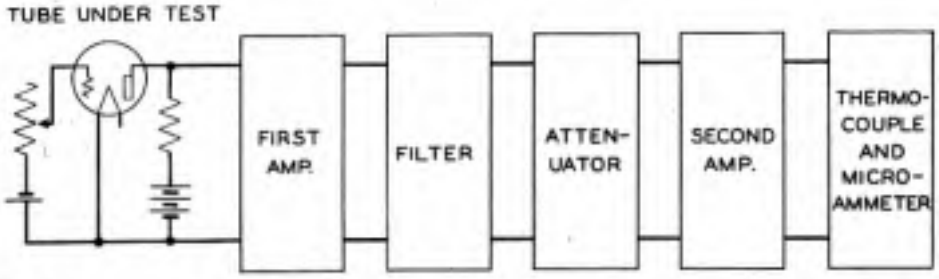


Fig. 2—Schematic amplifier circuit for measuring fluctuation noise in vacuum tubes.

The input circuit consists of the tube under test together with the variable grid resistor, external load resistor, and batteries for furnishing the required filament, grid, and plate voltages. Because of the high value of amplification required and the wide frequency range covered by the amplifier, this circuit required shielding from external disturbances arising from electrical, mechanical, and acoustical shock. Accordingly the tube under test was suspended by means of rubber bands, the whole circuit with the exception of batteries placed inside a tightly sealed lead lined box, and this box in turn suspended by means of a system of damped springs. The box with its cover removed and the tube in place is shown in Fig. 3. This shielding was sufficient to reduce the noise from outside disturbances to such a low level that no correction had to be made for it at any time.

The high gain amplifier [12] consists of two separate resistance coupled units each containing three stages. Each unit is so designed and shielded that the effect of external disturbances is eliminated. The total gain obtainable is about 165 db (constant to within 2 db from 10 cycles to 15,000 cycles). Since this gain is in excess of that required for the study of thermal and tube noises, an attenuator having a range of 63 db was inserted between the two units. In order to limit amplification to certain desired frequency bands, specially designed electric filters were inserted between the first amplifier unit and the attenuator. Three such filters were used of which one is a low-pass filter with cut-off around 205 cycles, and the other two are band-pass filters with mid-frequencies at 1750 and 11,000 cycles respectively. The frequency characteristic of the amplifier with no filter and with each filter inserted is shown in Fig. 4.

The recording instrument is a 600-ohm vacuum thermocouple and microammeter. Conveniently, the deflection of the microammeter is closely proportional to the mean square voltage applied to the couple. The procedure in making a measurement of tube noise is as follows:

[12] The essential parts of this amplifier were designed by Mr. E. T. Burton.

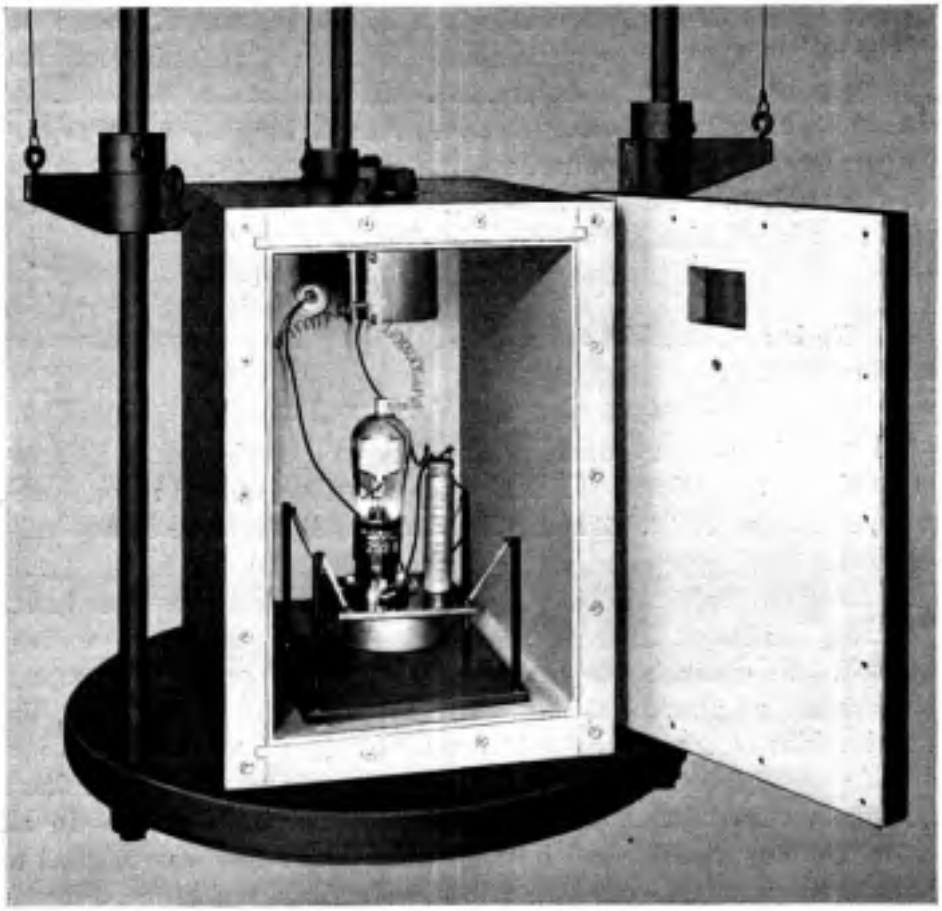


Fig. 3—Tube under test mounted in the shielding box.

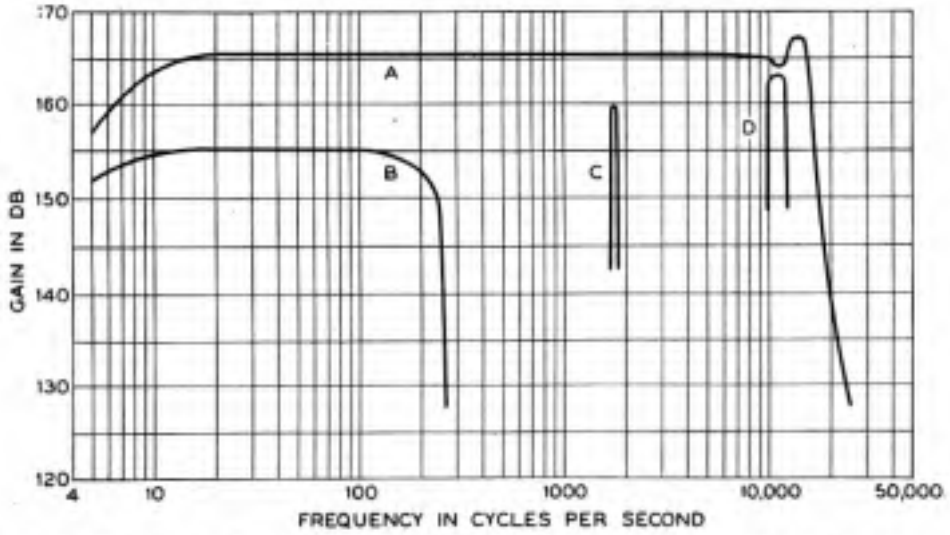


Fig. 4—Frequency characteristic of amplifier circuit. Curve *A* with no filter, Curve *B* with low pass filter, and Curves *C* and *D* with band pass filters.

With the tube under test operating at zero grid resistance, the attenuator is adjusted to give a convenient deflection of the microammeter (due to noise in the tube under test). Grid resistance is now added until this deflection is exactly doubled, thus making $\overline{E_N^2}$ equal to $\overline{E_T^2}$. This value of input resistance, designated by R_G, is a measure of the inherent noise of the tube. Substituting R_G in equation (13) the tube noise is calculated from the relation

$$V_F^2 = 4kTR_G = 1.64 \times 10^{-20} R_G \text{ volt}^2, \tag{14}$$

where R_G is expressed in ohms and T is 300° K. (approximate room temperature).

Noise in Certain Vacuum Tubes[13]

Quantitative measurements of tube noise were made on four different types of standard Western Electric vacuum tubes, namely: Nos. 102G, 264B, 262A and 259B. These tubes have as low a noise as any tube obtainable at the present time.

In order to obtain the best signal to noise ratios it was found that operating conditions different from those normally recommended must be used. In general, the cathode must be operated at as high a temperature as possible without impairing the life of the tube, the negative bias of the control grid must be reduced to as near zero as possible without causing excessive grid current, and the plate voltage must be reduced below the value normally recommended. In all the measurements described here the tube under test was coupled to the first amplifier unit through a 50,000-ohm load resistance. It was found that the signal-to-noise ratio could be improved a fraction of a db by increasing the load resistance (in accordance with equation (4)); this, however, necessitated a large plate voltage which was inconvenient. Six tubes of each type were tested and the noise data given below were obtained by averaging the six measurements for each type. Individual tubes may differ from these average values by as much as ± 1 db.

No. 102G Tube

This is a three-element, filament-type tube. Its long life, exceptionally high stability of operation, and good temperature saturation make it a desirable tube to use in the input stage of certain high-gain amplifiers. This tube also has a comparatively small microphonic response to mechanical and acoustical shock although it is not as good as the No. 262A and the No. 264B tubes in this respect.[14]

[13] Noise in other types of vacuum tubes has been reported by G. F. Metcalf and T. M. Dickinson, *Physics, 3*, 11 (1932); E. A. Johnson and C. Neitzert, *Rev. Sc. Inst., 5*, 196 (1934); E. B. Moullin and H. D. M. Ellis, *I. E. E. Jour., 74*, 323 (1934); W. Brentzinger and H. Viehmann, *Arch. f. Hochfr. und Elektroauk, 39*, 199 (1932).

[14] The microphonic response of several types of Western Electric vacuum tubes to mechanical agitation is reported by D. B. Penick in this issue of the *Bell Sys. Tech. Jour.*

The conditions found most suitable for quiet operation of the No. 102G tube and the corresponding average tube characteristics are given in the first two columns of Table I. Under these conditions the

TABLE I

WESTERN ELECTRIC NO. 102G TUBE

Operating Conditions	Tube Characteristics	Noise Data		
		Frequency Range Cycles per Sec.	V_F^2 Volt²	R_G Ohms
Filament Voltage, 2.0 volts Current, 1.0 ampere	Type of Tube, 3 Element Type of Cathode, Oxide Coated Filament			
Grid Voltage, − 0.5 volt		10–15,000	0.64×10^{-16}	3,900
Plate Voltage, 130 volts Current, 1.2 milliamperes	Amplification Factor, 30	5–205	2.2	13,600
	Plate Resistance, 45,000 ohms	1,750–1,850	0.58	3,550
Load Resistance, 50,000 ohms	Approx. Dynamic Input Capacitance, 80 $\mu\mu$f	10,000–12,000	0.54	3,300

average equivalent tube noise voltage, referred to the grid circuit, is given in the last column of the same table. These noise data are given in terms of R_G, the experimentally determined equivalent noise resistance of the tube, and in terms of V_F^2, calculated by means of equation (14), for each of the four frequency ranges shown in Fig. 4.

The No. 102G has the lowest noise of all the tubes tested and was found suitable for use in the first stage of high-gain amplifiers where tube noise is the limiting factor, provided it is not required that the input capacitance and microphonic response to mechanical and acoustical shock be extremely low.

No. 264B Tube

This is a three-element filament-type tube. Due to the rigid construction and the short filament which is designed to reduce vibration to a minimum, the microphonic response of the tube to mechanical and acoustical shock is exceptionally low.[15] The extensive system of spring suspensions and the heavy sound-proof chamber usually required for shielding low noise tubes may be simplified when using the No. 264B. In addition, this tube has good temperature saturation, low power consumption, and high stability of operation.

The operating conditions and noise data for this tube are given in Table II. Although the noise of this tube is slightly higher than that

[15] M. J. Kelly, *S. M. P. E. Jour.*, **18**, 761 (1932).

TABLE II

WESTERN ELECTRIC No. 264B TUBE

Operating Conditions	Tube Characteristics	Noise Data		
		Frequency Range Cycles per Sec.	Vp^2 Volt^2	Rg Ohms
Filament Voltage, 1.5 volts Current, .30 ampere	Type of Tube, 3 Element Type of Cathode, Oxide Coated Filament			
Grid Voltage, — 0.5 volt	Amplification Factor, 7	10–15,000 5–205	1.3×10^{-16} 6.6	7,650 40,000
Plate Voltage, 26 volts Current, 0.6 milliampere	Plate Resistance, 18,500 ohms	1,750–1,850	1.1	6,800
Load Resistance, 50,000 ohms	Approx. Dynamic Input Capacitance, 30 $\mu\mu$f	10,000–12,000	1.0	6,200

of the No. 102G, the lower microphonic response and the lower power consumption make it a more desirable tube to use in input stages of certain high gain amplifiers.

No. 262A Tube

This is a three-element tube having an indirectly heated cathode. It is designed to give a microphonic response to mechanical and acoustical shock [15] still lower than that of the 264B. Except for frequencies below 200 cycles per second it was found that no acoustic shield was necessary for this tube even when working at extremely low levels. Although this tube is designed to have a low hum disturbance resulting from alternating current for heating the cathode (the interference from this effect can be held to less than 7×10^{-6} equivalent input volt), direct current power was used in the measurements here described.

The operating conditions and noise data for the No. 262A tube are given in Table III.

No. 259B Tube

This is a four-element, screen-grid tube having an indirectly heated cathode. Its comparatively high amplification factor makes possible a relatively large gain per stage so that when it is used in the first stage of a high-gain amplifier succeeding stages contribute nothing to the total noise.

Noise measurements on the No. 259B tube show that the signal-to-noise ratio is approximately independent of the plate voltage over a

TABLE III

WESTERN ELECTRIC NO. 262A TUBE

Operating Conditions	Tube Characteristics	Noise Data		
		Frequency Range Cycles per Sec.	VF^2 Volt²	R_G Ohms
Heater Voltage, 10 volts Current, 0.32 ampere	Type of Tube, 3 Element Type of Cathode, Oxide Coated, Indirectly Heated	10–15,000	1.3×10^{-16}	7,700
Grid Voltage, − 1.0 volt	Amplification Factor, 15.7	5–205	17.	100,000
Plate Voltage, 44 volts Current, 1.0 milliampere	Plate Resistance, 22,000 ohms	1,750–1,850	1.0	6,400
Load Resistance, 50,000 ohms	Approx. Dynamic Input Capacitance, 23 $\mu\mu$f	10,000–12,000	0.84	5,100

wide operating range, but is closely dependent on the plate current as affected by the control and screen grid voltages. Table IV contains the operating conditions and noise data for this tube.

Noise measurements were also made on the No. 259B tube with its control grid floating at equilibrium potential. Using the operating voltages specified above, the noise level was about 20 db higher than those given in Table IV. The level can be greatly reduced by operating the tube at a lower cathode temperature and with lower screen

TABLE IV

WESTERN ELECTRIC NO. 259B TUBE

Operating Conditions	Tube Characteristics	Noise Data		
		Frequency Range Cycles per Sec.	VF^2 Volt²	R_G Ohms
Heater Voltage, 2.0 volts Current, 1.7 amperes	Type of Tube, 4 Element Screen Grid Type of Cathode, Oxide Coated, Indirectly Heated	10–15,000	3.2×10^{-16}	19,800
Grid Control Voltage, − 1.5 volts	Amplification Factor, 1,500	5–205	7.7	47,000
Screen Voltage, 22.5 volts	Plate Resistance, 2.75 megohms	1,750–1,850	2.8	17,100
Plate Voltage, 100 volts Current, 0.6 milliampere	Approx. Dynamic Input Capacitance, 6.0 $\mu\mu$f	10,000–12,000	2.8	17,000
Load Resistance, 50,000 ohms				

and plate voltages.[16] This reduction in noise is due to a decrease in current to the floating grid. Using a heater current of 1.3 amperes, a plate current of 0.1 milliampere, a screen potential of 16.5 volts and a plate potential of 30 volts the equivalent input noise was 1.4×10^{-5} volt for the entire frequency range from 10 cycles to 15,000 cycles. Under these operating conditions the floating grid potential was 1.0 volt negative with respect to the cathode, the input resistance 1.4×10^{10} ohms, the dynamic grid-to-cathode capacitance 6×10^{-12} farad, and each component of grid current about 4.5×10^{-12} ampere.

Discussion of Results

From the noise data in the preceding tables one can estimate quite accurately the equivalent input noise voltage of each of the four types of tubes at any frequency between 5 and 15,000 cycles, and for any band width within these limits. For example, using the noise data given in Table I the equivalent input noise voltage of the No. 102G tube working over a band having sharp cut-offs at 5 cycles and 205 cycles is computed to be

$$(\overline{V^2})^{1/2} = (V_F^2 F)^{1/2} = 2.1 \times 10^{-7} \text{ volt.} \qquad (15)$$

For a band width of 200 cycles with mid-frequency at 10,000 cycles this noise is reduced to 1.0×10^{-7} volt. It can be seen that for each type of tube the noise voltage over equal band widths is between 1.5 and 4.5 times greater at frequencies below 200 cycles than at the higher frequencies.[17]

Even at high frequencies the noise voltage is above that expected from thermal noise in the plate circuit which, as stated above, is the absolute minimum to which fluctuation noise in a thermionic vacuum tube may be reduced after all other causes are eliminated. In the case of the No. 102G tubes for instance, using the operating conditions of Table I, and assuming 1100° K. as the temperature of the barium oxide filament, it is found by means of equation (4) that the equivalent input noise voltage produced by thermal agitation in the plate circuit is 2.7×10^{-8} volt for a band width of 200 cycles. The total input noise voltage obtained experimentally at the higher frequencies is greater than this by a factor of 3.8. In like manner it is found that the total input noise voltages found experimentally for the Nos. 264B, 262A and 259B tubes are greater than the equivalent input thermal

[16] I am indebted to Dr. J. R. Dunning of Columbia University for pointing out this fact.

[17] Other investigators have also found an increase in tube noise energy at the lower frequencies. G. F. Metcalf and T. M. Dickinson, *Physics*, 3, 11 (1932).

noise voltages produced in the plate circuit by factors 2.1, 3.7, and 16 respectively. These calculations show that each of these four types of tubes approaches the requirements of an ideal low noise amplifying tube although none of them is perfect in this respect.

As stated above, the best signal-to-noise ratio in a high-gain amplifier is obtained when thermal agitation in the input resistance is responsible for most of the noise in the amplifier. This condition is met when the resistance of the input circuit is higher than the value of R_G for the input tube. In case the resistance in the input circuit is less than R_G the input signal and the thermal noise from the input circuit can be raised above the noise of the tube by using an input transformer having a sufficiently high voltage step-up. The voltage ratio of the transformer, and in turn the possible ratio of input circuit thermal noise to tube noise, is limited, especially at the higher frequencies, by the dynamic grid-to-ground capacitance of the input tube and its leads. In such a circuit the No. 259B tube with its lower inter-electrode capacities and higher tube noise is often more desirable than even the quietest three-element tubes.

In those high-gain amplifiers in which unavoidably the resistance of the input circuit is low, the tube rather than thermal agitation in the grid circuit is responsible for most of the noise. Here the best signal-to-noise ratio can be obtained by choosing a tube for the initial stage having the lowest possible noise level. The above measurements show that one of the three-element tubes, particularly the No. 102G tube if sufficient shielding is used, is best suited for this purpose.

The lower limits of noise obtainable with high gain amplifiers may be estimated by means of Fig. 5, which shows the noise as a function of input resistance and frequency band width when thermal agitation in the input circuit is responsible for all the noise. The data for this figure are obtained from the thermal noise relationship

$$\overline{V_T^2} = 1.64 \times 10^{-20} RF \text{ volt}^2. \tag{16}$$

R is expressed in ohms and the temperature has been taken at 300° K., which is approximately room temperature. It must be remembered that the attainment of these noise levels at low input resistances is limited by the input transformer.

The results of the noise measurements on the No. 259B tube with floating grid may be compared with the value predicted by equations (8) and (9). Inserting the tube characteristics obtained by experiment ($r_g = 1.4 \times 10^{10}$ ohms, $i_g = 9 \times 10^{-12}$ ampere, and $c = 6 \times 10^{-12}$ farad), and integrating between the frequency limits 10 cycles

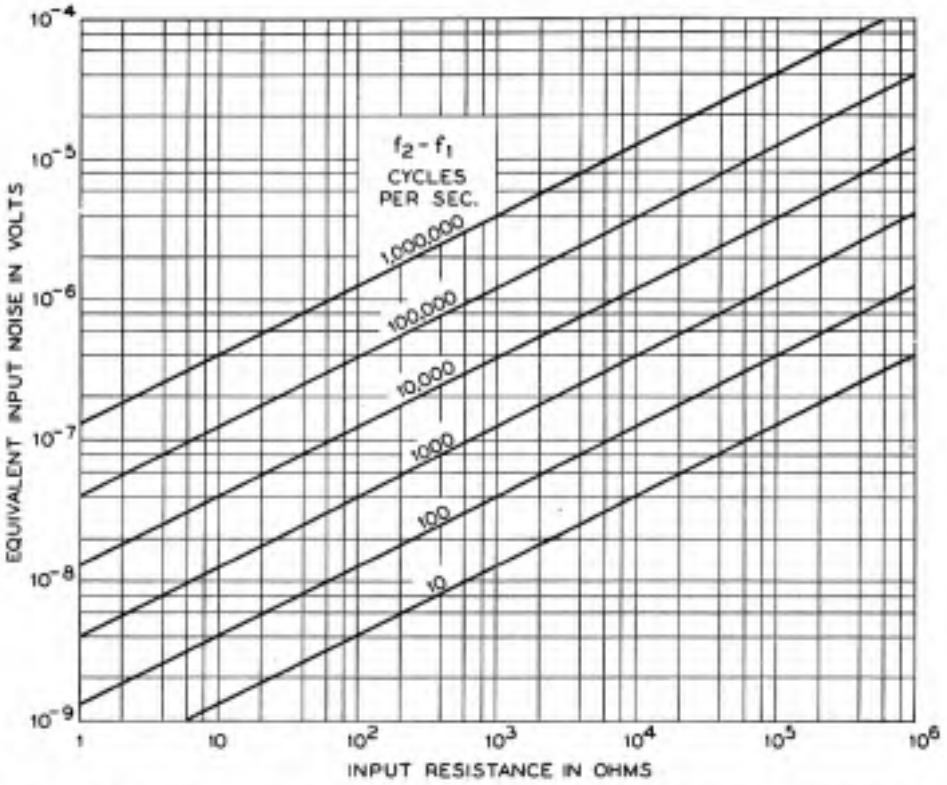


Fig. 5—Thermal noise level as a function of input resistance and frequency range.

and 15,000 cycles, it is found that the equivalent thermal noise input is 0.9×10^{-5} volt, while the shot noise input is 1.4×10^{-5} volt. The total noise is the square root of the sum of the squares of these values or 1.7×10^{-5} volt. This agrees with the measured value of 1.4×10^{-5} volt within an error of 20 per cent, which is as accurate as the determination of the grid currents. These equations may also be used to calculate the noise originating in the grid circuit when external resistance or capacitance is connected between grid and cathode, provided r_g and c are now calculated from the internal and external impedances in parallel.

A common method of detecting corpuscular or electromagnetic radiation makes use of an ionization chamber and linear amplifier. In this circuit the control grid in the first tube of the amplifier is connected to the collecting electrode of the ionization chamber and both allowed to float at equilibrium potential.[18] The shot and thermal noise in this grid circuit sets a limit to the measurement of extremely weak radiation. Knowing the value of input capacitance, input

[18] H. Greinacher, *Zeits. f. Physik*, **36**, 364 (1926).

resistance, floating grid current, and the frequency limits of the amplifier, equations (8) and (9) may be used to calculate this limiting noise level. For example, if one uses a No. 259B tube with the operating voltages specified for floating grid, an ionization chamber having a capacitance of 15×10^{-12} farad, and an amplifier having a frequency range from 200 to 5000 cycles per second the limiting noise level is 1×10^{-6} root mean square volt.

The limiting noise level in a system consisting of a photoelectric cell and thermionic amplifier is determined by thermal agitation in the coupling circuit between the photoelectric cell and amplifier, and by shot noise in the photoelectric current (in circuits where the photoelectric current is very small and the coupling resistance is very high, shot noise from grid current in the vacuum tube becomes appreciable). The noise of thermal agitation may be calculated by means of equation (8) provided r_g is now replaced by R, the coupling resistance. If vacuum cells are used, the photoelectric current produces a pure shot noise which can be calculated by equation (9) provided i_g is replaced by I, the photoelectric current. In gas filled photocells where collision ionization occurs, the noise is in excess of the value calculated in this manner.[19] The relative magnitude of shot noise and thermal noise depends on the values of I and R, and by combining equations (8) and (9) it is found that

$$\overline{V_S^2}/\overline{V_T^2} = eIR/2kT = 19.4IR, \tag{17}$$

where I is expressed in amperes, R in ohms, and T is 300° K. Thus an increase in either I or R will tend to make shot noise exceed thermal noise. This is the desirable condition since it furnishes the largest ratio of signal-to-noise for a given light signal on the photoelectric cell.

In conclusion I wish to acknowledge my indebtedness to Dr. J. B. Johnson for the helpful criticism he has given during the course of this work.

[19] B. A. Kingsbury, *Phys. Rev.*, **38**, 1458 (1931).

Systems for Wide-Band Transmission Over Coaxial Lines

By L. ESPENSCHIED and M. E. STRIEBY

In this paper systems are described whereby frequency band widths of the order of 1000 kc. or more may be transmitted for long distances over coaxial lines and utilized for purposes of multiplex telephony or television. A coaxial line is a metal tube surrounding a central conductor and separated from it by insulating supports.

IT appears from recent development work that under some conditions it will be economically advantageous to make use of considerably wider frequency ranges for telephone and telegraph transmission than are now in use [1,2] or than are covered in the recent paper on carrier in cable.[3] Furthermore, the possibilities of television have come into active consideration and it is realized that a band of the order of one million cycles or more in width would be essential for television of reasonably high definition if that art were to come into practical use.[4,5]

This paper describes certain apparatus and structures which have been developed to employ such wide frequency ranges. The future commercial application of these systems will depend upon a great many factors, including the demand for additional large groups of communication facilities or of facilities for television. Their practical introduction is, therefore, not immediately contemplated and, in any event, will necessarily be a very gradual process.

Types of High-Frequency Circuits

The existing types of wire circuits can be worked to frequencies of tens of thousands of cycles, as is evidenced by the widespread application of carrier systems to the open-wire telephone plant and by the development of carrier systems for telephone cable circuits.[2,3] Further development may lead to the operation of still higher frequencies over the existing types of plant. However, for protection against external interference these circuits rely upon balance, and as the frequency band is widened, it becomes more and more difficult to maintain a sufficiently high degree of balance. The balance requirements may be made less severe by using an individual shield around

[1] For references, see end of paper.

* Published in *Electrical Engineering*, October, 1934. Scheduled for presentation at Winter Convention of A. I. E. E., New York, N. Y., January, 1935.

each circuit, and with sufficient shielding balance may be entirely dispensed with.

A form of circuit which differs from existing types in that it is unbalanced (one of the conductors being grounded), is the coaxial or concentric circuit. This consists essentially of an outer conducting tube which envelops a centrally-disposed conductor. The high-frequency transmission circuit is formed between the inner surface of the outer conductor and the outer surface of the inner conductor. Unduly large losses at the higher frequencies are prevented by the nature of the construction, the inner conductor being so supported within the tube that the intervening dielectric is largely gaseous, the separation between the conductors being substantial, and the outer conductor presenting a relatively large surface. By virtue of skin effect, the outer tube serves both as a conductor and a shield, the desired currents concentrating on its inner surface and the undesired interfering currents on the outer surface. Thus, the same skin effect which increases the losses within the conductors provides the shielding which protects the transmission path from outside influences, this protection being more effective the higher the frequency.

The system which this paper outlines has been based primarily upon the use of the coaxial line. The repeater and terminal apparatus described, however, are generally applicable to any type of line, either balanced or unbalanced, which is capable of transmitting the frequency range desired.

The Coaxial System

A general picture of the type of wide band transmission system which is to be discussed is briefly as follows: A coaxial line about 1/2 inch in outside diameter is used to transmit a frequency band of about 1,000,000 cycles, with repeaters capable of handling the entire band placed at intervals of about 10 miles. Terminal apparatus may be provided which will enable this band either to be subdivided into more than 200 telephone circuits or to be used *en bloc* for television.

Such a wide-band system is illustrated in Fig. 1. It is shown to comprise several portions, namely, the line sections, the repeaters, and the terminal apparatus, the latter being indicated in this case as for multiplex telephony. Two-way operation is secured by using two lines, one for either direction. It would be possible, however, to divide the frequency band and use different parts for transmission in opposite directions.

A form of flexible line which has been found convenient in the experimental work is illustrated in Fig. 2 and will be described more fully

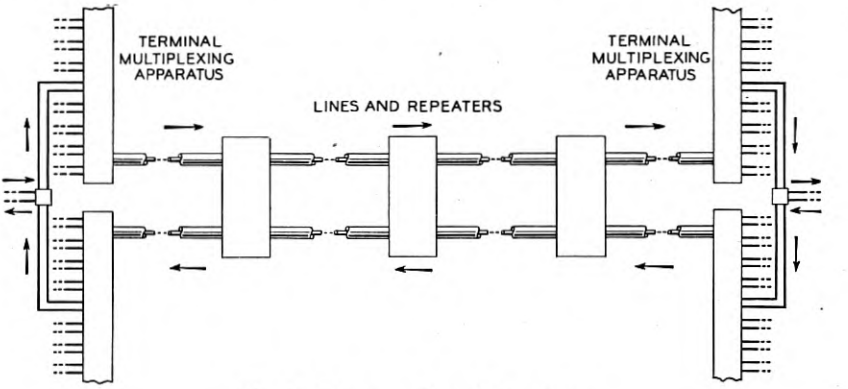


Fig. 1—Diagram of coaxial system.

subsequently. Such a coaxial line can be constructed to have the same degree of mechanical flexibility as the familiar telephone cable. While this line has a relatively high loss at high frequencies, the transmission path is particularly well adapted to the frequent application of repeaters, since the shielding permits the transmission currents to fall to low power levels at the high frequencies.

Of no little importance also is the fact that the attenuation-frequency characteristic is smooth throughout the entire band and obeys a simple law of change with temperature. (This is due to the fact that the dielectric is largely gaseous and that insulation material of good dielectric properties is employed.) This smooth relation is extremely

Fig. 2—Small flexible coaxial structure.

helpful in the provision of means in the repeaters for automatically compensating for the variations which occur in the line attenuation with changes of temperature. This type of system is featured by large transmission losses which are offset by large amplification, and it is necessary that the two effects match each other accurately at all times throughout the frequency range.

It will be evident that the repeater is of outstanding importance in this type of system, for it must not only transmit the wide band of frequencies with a transmission characteristic inverse to that of the line, with automatic regulation to care for temperature changes, but must also have sufficient freedom from inter-modulation effects to permit the use of large numbers of repeaters in tandem without objec-

tionable interference. Fortunately, recent advances in repeater technique have made this result possible, as will be appreciated from the subsequent description.

An interesting characteristic of this type of system is the way in which the width of the transmitted band is controlled by the repeater spacing and line size, as follows:

1. *For a given size of conductor and given length of line, the band width increases nearly as the square of the number of the repeater points.* Thus, for a coaxial circuit with about .3-inch inner diameter of outer conductor, a 20-mile repeater spacing will enable a band up to about 250,000 cycles to be transmitted, a 10-mile spacing will increase the band to about 1,000,000 cycles, and a 5-mile spacing to about 4,000,000 cycles.
2. *For a given repeater spacing, the band width increases approximately as the square of the conductor diameter.* Thus, whereas a tube of .3-inch inner diameter will transmit a band of about 1,000,000 cycles, .6-inch diameter will transmit about 4,000,000 cycles, while a diameter corresponding to a full-sized telephone cable might transmit something of the order of 50,000,000 cycles, depending upon the dielectric employed and upon the ability to provide suitable repeaters.

Earlier Work

It may be of interest to note that as a structure, the coaxial form of line is old—in fact, classical. During the latter half of the last century it was the object of theoretical study, in respect to skin effect and other problems, by some of the most prominent mathematical physicists of the time. Reference to some of this work is made in a paper by Schelkunoff, dealing with the theory of the coaxial circuit.[6]

On the practical side, it is found on looking back over the art that the coaxial form of line structure has been used in two rather widely different applications: first, as a long line for the transmission of low frequencies, examples of which are usage for submarine cables,[7, 8] and for power distribution purposes, and second as a short-distance, high-frequency line serving as an antenna lead-in.[9, 10]

The coaxial conductor system herein described may be regarded as an extension of these earlier applications to the long-distance transmission of a very wide range of frequencies suitable for multiplex telephony or television.[11] Although dealing with radio frequencies, this system represents an extreme departure from radio systems in that a relatively broad band of waves is transmitted, this band being con-

fined to a small physical channel which is shielded from outside disturbances. The system, in effect, comprehends a frequency spectrum of its own and shuts it off from its surroundings so that it may be used again and again in different systems without interference.

This new type of facility has not yet been commercially applied. It is, in fact, still in the development stage. Sufficient progress has already been made, however, to give reasonable assurance of a satisfactory solution of the technical problems involved. This progress is outlined below under three general headings: (1) the coaxial line and its transmission properties, (2) the wide band repeaters, and (3) the terminal apparatus.

The Coaxial Line

An Experimental Verification

One of the first steps taken in the present development was in the nature of an experimental check of the coaxial conductor line, designed primarily to determine whether the desirable transmission properties which had been disclosed by a theoretical study could be fully realized under practical conditions. For this purpose a length of coaxial structure capable of accurate computation was installed near Phoenixville, Pa. Figure 3 shows a sketch of the structure used and gives its dimensions. It comprised a copper tube of 2.5 inches outside diameter, within which was mounted a smaller tube which, in turn, contained a small copper wire. Two coaxial circuits of different sizes were thus made available, one between the outer and the inner tubes, and the other between the inner tube and the central wire. The installation comprised two 2600-foot lengths of this structure.

The diameters of these coaxial conductors were so chosen as to obtain for each of the two transmission paths a diameter ratio which approximates the optimum value, as discussed later. The conductors were separated by small insulators of isolantite. The rigid construction and the substantial clearances between conductors made it possible to space the insulators at fairly wide intervals, so that the dielectric between conductors was almost entirely air. The outer conductor was made gas-tight, and the structure was dried out by circulating dry nitrogen gas through it. The two triple conductor lines were suspended on wooden fixtures and the ends brought into a test house, as shown in Fig. 4.

The attenuation was measured by different methods over the frequency range from about 100 kilocycles to 10,000 kilocycles. Investigation showed that the departures from ideal construction occasioned by the joints, the lack of perfect concentricity, etc., had remark-

ably little effect on the attenuation. In order to study the effect of eccentricity upon the attenuation, tests were made in which this effect was much exaggerated, and the results substantiated theoretical predictions. The impedance of the circuits was measured over the same range as the attenuation. A few measurements on a short length were made at frequencies as high as 20,000 kilocycles.

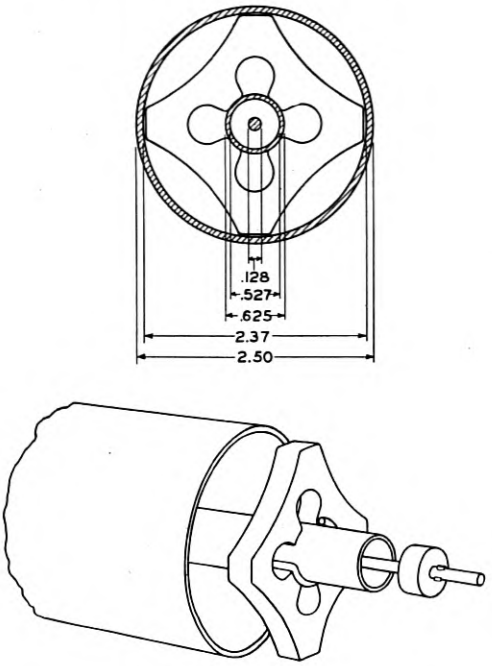


Fig. 3—Structure used in Phoenixville installation.

Measurements were secured of the shielding effect of the outer conductor of the coaxial circuit up to frequencies in the order of 100 to 150 kilocycles, the results agreeing closely with the theoretical values. Above these frequencies, even with interfering sources much more powerful than would be encountered in practice, the induced currents dropped below the level of the noise due to thermal agitation of electricity in the conductors (resistance noise) and could not be measured.

The preliminary tests at Phoenixville, therefore, demonstrated that

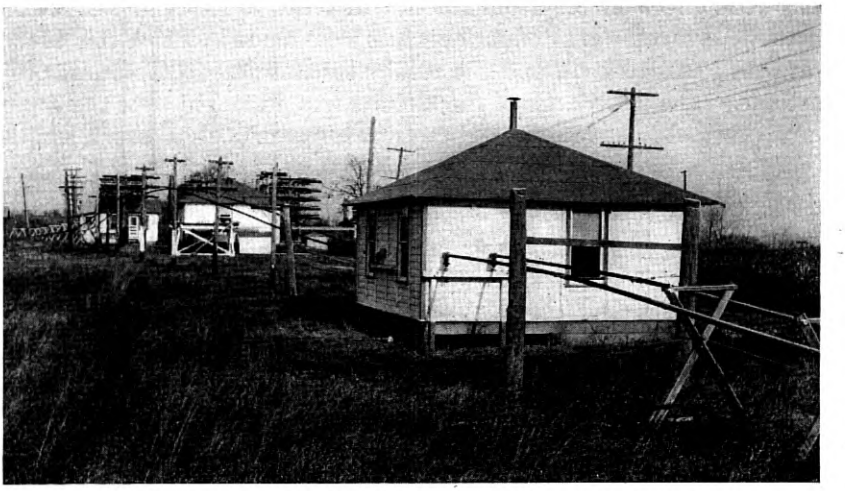


Fig. 4—Phoenixville installation showing conductors entering test house.

a practical coaxial circuit, with its inevitable mechanical departures from the ideal, showed transmission properties substantially in agreement with the theoretical predictions.

Small Flexible Structures

Development work on wide-band amplifiers, as discussed later, indicated the practicability of employing repeaters at fairly close intervals. This pointed toward the desirability of using sizes of coaxial circuit somewhat smaller than the smaller of those used in the preliminary experiments, and having correspondingly greater attenuation. Furthermore, it was desired to secure flexible structures which could be handled on reels after the fashion of ordinary cable. Accordingly, several types of flexible construction, ranging in outer diameter from about .3 inch to .6 inch, have been experimented with. Structures were desired which would be mechanically and electrically satisfactory, and which could be manufactured economically, preferably with a continuous process of fabrication.

One type of small flexible structure which has been developed is shown in Fig. 2. The outer conductor is formed of overlapping copper strips held in place with a binding of iron or brass tape. The insulation consists of a cotton string wound spirally around the inner conductor, which is a solid copper wire. This structure has been made in several sizes of the order of 1/2 inch diameter or less. When it is to be used as an individual cable, the outer conductor is surrounded by a

lead sheath, as shown, to prevent the entrance of moisture. One or more of the copper tape structures without individual lead sheath may be placed with balanced pairs inside a common cable sheath.

Another flexible structure is shown in Fig. 5. The outer conductor in this case is a lead sheath which directly surrounds the inner conductor with its insulation. Since lead is a poorer conductor than copper, it is necessary to use a somewhat larger diameter with this construction in order to obtain the same transmission efficiency. Lead is also inferior to copper in its shielding properties and to obtain the same degree of shielding the lead tube of Fig. 5 must be made correspondingly thicker than is necessary for a copper tube.

The insulation used in the structure shown in Fig. 5 consists of hard

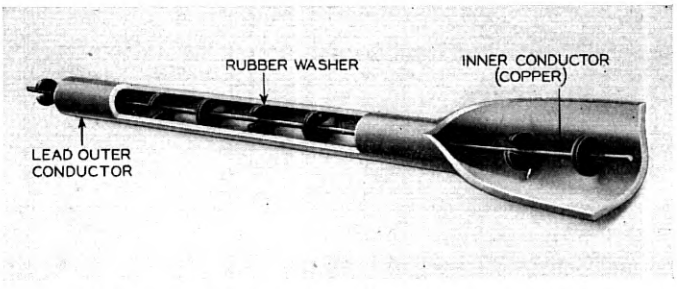


Fig. 5—Coaxial structure with rubber disc insulators.

rubber discs spaced at intervals along the inner wire. Cotton string or rubber disc insulation may be used with either form of outer tube. The hard rubber gives somewhat lower attenuation, particularly at the higher frequencies.

Another simple form of structure employs commercial copper tubing into which the inner wire with its insulation is pulled. Although this form does not lend itself readily to a continuous manufacturing process, it may be advantageous in some cases.

Transmission Characteristics

Attenuation

At high frequencies the attenuation of the coaxial circuit is given closely by the well-known formula:

$$\alpha = \frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}}, \tag{1}$$

where R, L, C and G are the four so-called "primary constants" of the line, namely, the resistance, inductance, capacitance and conductance

per unit of length. The first term of (1) represents the losses in the conductors, while the second term represents those in the dielectric.

When the dielectric losses are small, the attenuation of a coaxial circuit increases, due to skin effect in the conductors, about in accordance with the square root of the frequency. With a fixed diameter ratio, the attenuation varies inversely with the diameter of the circuit. By combining these relations there are obtained the laws of variation of band width in accordance with the repeater spacing and the size of circuit, as stated previously.

The attenuation-frequency characteristic of the flexible structure illustrated in Fig. 2, with about .3 inch diameter, is given in Fig. 6.

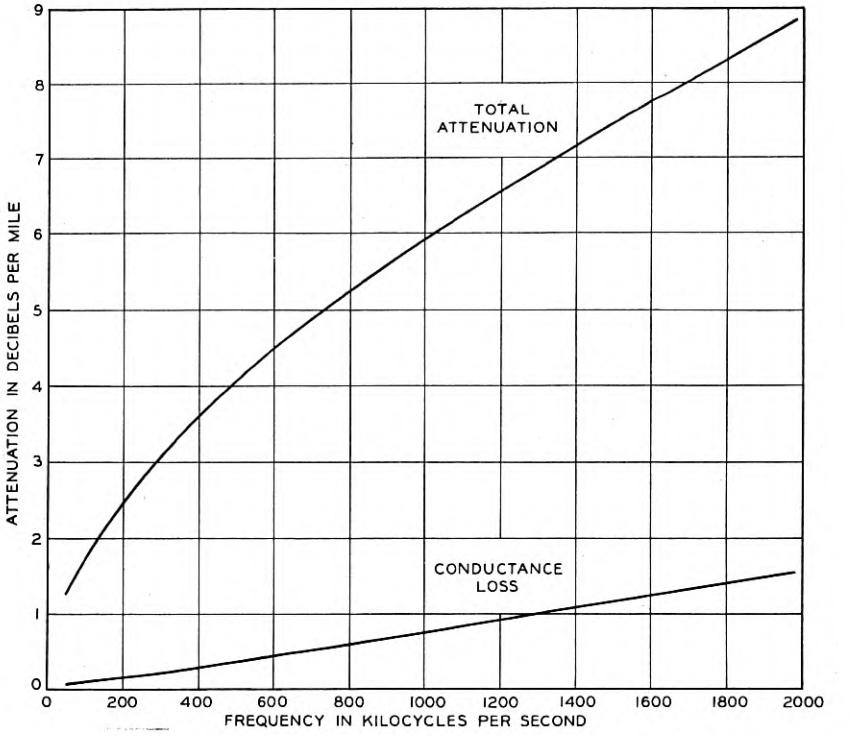


Fig. 6—Attenuation of small flexible coaxial structure (Fig. 2).

The figure shows also that the conductance loss due to the insulation is a small part of the total.

It is interesting to compare the curves of the transmission characteristics of the coaxial circuit with those of other types of circuits. Figure

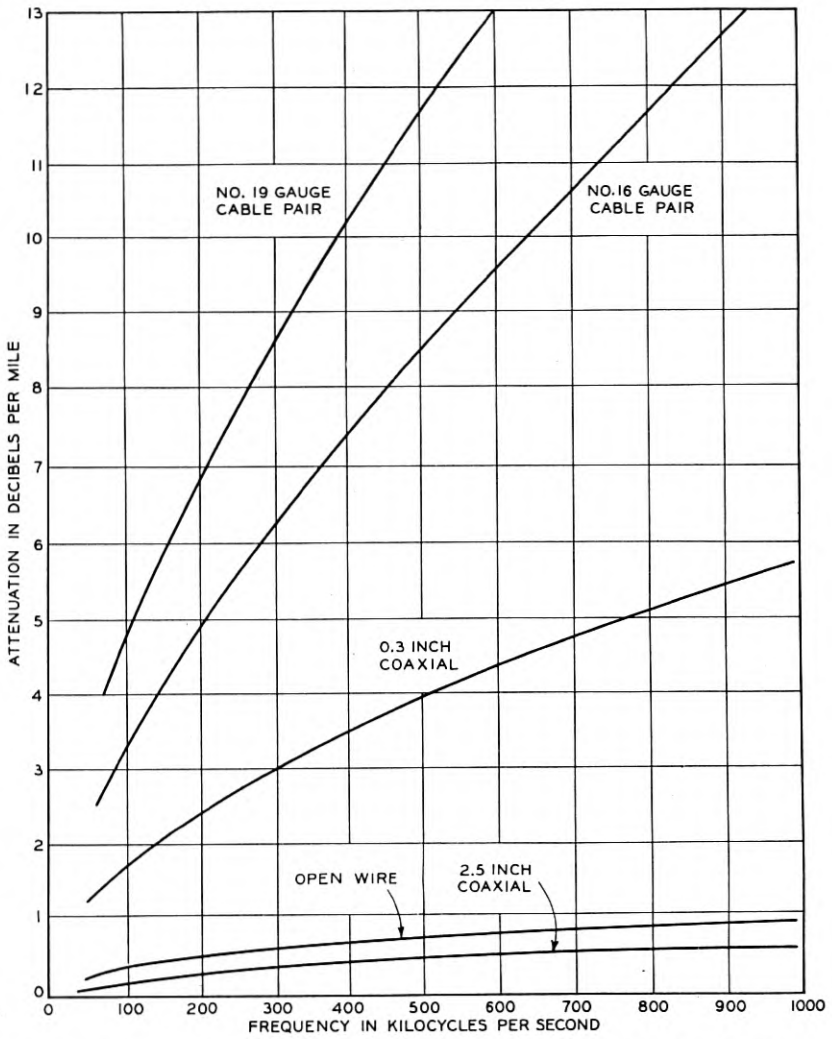


Fig. 7—Attenuation frequency characteristics of coaxial and other circuits.

7 shows the high-frequency attenuation of two sizes of coaxial circuit using copper tube outer conductors, of .3 inch and 2.5 inch inner diameter, and that of cable and open-wire pairs in the same frequency range.

Effect of Eccentricity

The small effect of lack of perfect coaxiality upon the attenuation of a coaxial circuit is illustrated by the curve of Fig. 8, which shows

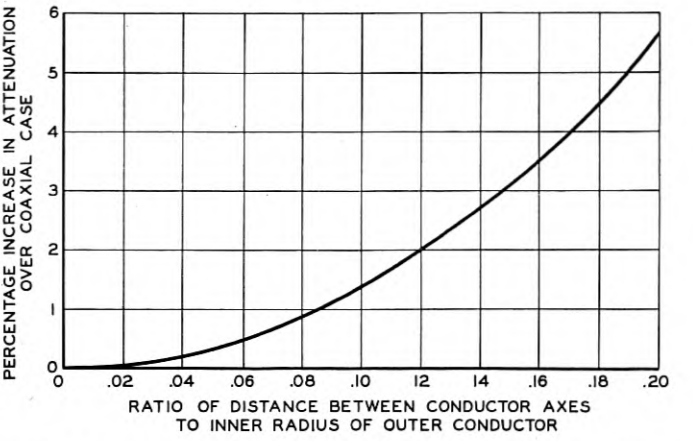


Fig. 8—Increase in attenuation of coaxial circuit due to eccentricity.

attenuation ratios plotted as a function of eccentricity, assuming a fixed ratio of conductor diameters and substantially air insulation.

Temperature Coefficient

With a coaxial circuit, as with other types of circuits, the temperature coefficient of resistance decreases as the frequency is increased, due to the action of skin effect, and approaches a value of one-half the d.-c. temperature coefficient.[12] Thus, for conductors of copper the a.-c. coefficient at high frequencies is approximately .002 per degree Centigrade. When the dielectric losses are small, the temperature coefficient of attenuation at high frequencies is the same as the temperature coefficient of resistance.

Diameter Ratio

An interesting condition exists with regard to the relative sizes of the two conductors. For a given size of outer conductor there is a unique ratio of inner diameter of outer conductor to outer diameter of inner conductor which gives a minimum attenuation. At high frequencies, this optimum ratio of diameters (or radii) is practically independent of frequency. When the conductivity is the same for both conductors, and either the dielectric losses are small or the insulation is distributed so that the dielectric flux follows radial lines, the value of the optimum diameter ratio is approximately 3.6. When the outer and inner conductors do not have the same conductivity, the optimum diameter ratio differs from this value. For a lead outer conductor and copper inner conductor, for example, the ratio should be about 5.3.

Stranding

Inasmuch as the resistance of the inner conductor contributes a large part of the high frequency attenuation of a coaxial circuit, it is natural to consider the possibility of reducing this resistance by employing a conductor composed of insulated strands suitably twisted or interwoven.[13] Experiments along this line showed that this method is impractical at frequencies above about 500 kilocycles, owing to the fineness of stranding required.

Characteristic Impedance

The high-frequency characteristic impedance of a coaxial circuit varies inversely with the square root of the effective dielectric constant, i.e., the ratio of the actual capacitance to the capacitance that would be obtained with air insulation. The impedance of a circuit having a given dielectric constant depends merely upon the ratio of conductor diameters and not upon the absolute dimensions. For a diameter ratio of 3.6, the impedance of a coaxial circuit with gaseous insulation is about 75 ohms.

Velocity of Propagation

For a coaxial circuit with substantially gaseous insulation, the velocity of propagation at high frequencies approaches the speed of light. Hence the circuit is capable of providing high velocity telephone channels with their well-recognized advantages. The fact that the velocity at high frequencies is substantially constant minimizes the correction required to bring the delay distortion within the limits required for a high quality television band.

Shielding and Crosstalk

The shielding effect of the outer conductor of a coaxial circuit is illustrated in Fig. 9, where the transfer impedance between the outer and inner surfaces of the outer conductor is plotted as a function of frequency. There will be observed the sharp decrease in inductive susceptibility as the frequency rises. On this account, the crosstalk between adjacent coaxial circuits falls off very rapidly with increasing frequency. The trend is, therefore, markedly different from that for ordinary non-shielded circuits which rely upon balance to limit the inductive coupling. As a practical matter, less shielding is ordinarily required to avoid crosstalk than to avoid external interference.

With suitable design the shielding effect of the outer conductor renders the coaxial circuit substantially immune to external interference at frequencies above the lower end of the spectrum. Hence the signals transmitted over the circuit may be permitted to drop

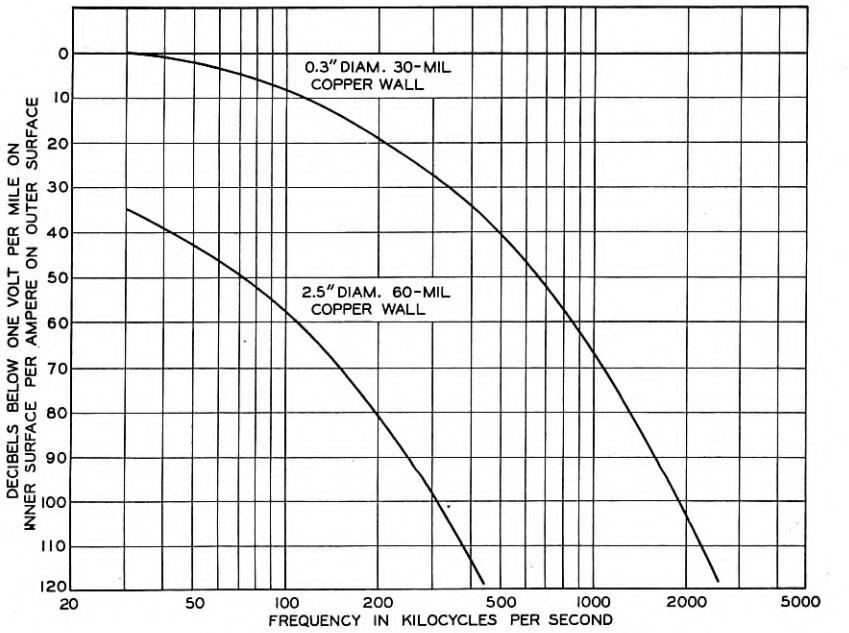


Fig. 9—Transfer impedance of coaxial circuit.

down to a level determined largely by the noise due to thermal agitation of electricity in the conductors and tube noise in the associated amplifiers. It appears uneconomical to make the outer conductor sufficiently thick to provide adequate shielding for the very low frequencies. Also it seems impractical to design the repeaters to transmit very low frequencies. Hence the best system design appears to be one in which the lowest five or ten per cent of the frequency range is not used for signal transmission. The coaxial circuit is, however, well suited to the transmission of 60-cycle current for operating repeaters, a matter which will be referred to later.

Broad-Band Amplifiers

In order to realize the full advantage of broad-band transmission, the repeater for this type of system should be capable of amplifying the entire frequency band *en bloc*. Furthermore, it should be so stable and free from distortion that a large number of repeaters may be operated in tandem. Although high-gain radio frequency amplifiers are in everyday use, these are generally arranged to amplify at any one time only a relatively narrow band of frequencies, a variable tuning device being provided so that the amplification may be obtained at

any point in a fairly wide frequency range. The high gain is usually obtained by presenting a high impedance to the input circuits of the various tubes through tuning the input and interstage coupling circuits to approximate anti-resonance.

In amplifying a broad band of frequencies, it is difficult to maintain a very high impedance facing the grid circuits. The inherent capacitances between the tube elements and in the mounting result in a rather low impedance shunt which can not be resonated over the desired frequency band. It is, therefore, necessary to use relatively low impedance coupling circuits and to obtain as high gain as possible from the tubes themselves. The amount of gain which can be obtained without regeneration depends, of course, upon the type of tube, the number of amplification stages, the band width, and also upon the ratio of highest to lowest frequency transmitted.

Repeater Gain

The total net gain desired in a line amplifier is such as to raise the level of an incoming signal from its minimum permissible value, which is limited by interference, up to the maximum value which the amplifier can handle.

As pointed out above, the noise in a well shielded system is that due to resistance noise in the line conductors and tube noise in the amplifiers. In some of the repeaters which have been built, the amplifier noise has been kept down to about 2 db above resistance noise, corresponding to about 7×10^{-17} watt per voice channel. In a long line with many repeaters the noise voltages add at random, or in other words, the noise powers add directly. Assuming, for example, a line with 200 repeaters, the noise power at the far end would be 200 times that for a single repeater section. In general, the line and amplifier noise will not be objectionable in a long telephone channel if the speech sideband level at any amplifier input is not permitted to drop more than about 55 db below the level of the voice frequency band at the transmitting toll switchboard.

The determination of the volume which a tube can handle in transmitting a wide band of frequencies involves a knowledge of the distribution in time and frequency of the signaling energy and of the requirements as to distortion of the various components of the signal. The distribution of the energy in telephone signals has been the subject of much study. This distribution is known to vary over very wide limits, depending upon the voice of the talker and many other factors. It is, therefore, obvious that the problem of summing up the energy of some hundreds of simultaneous telephone conversations is a difficult one.

Enough work has been done, however, to indicate fairly well what the result of such addition will be.

As to distortion in telephone transmission, the most serious problem has been to limit the intermodulation between various signals which are transmitted simultaneously through the repeater and appear as noise in the telephone channel. The requirement for such noise is similar to that for line and tube noise, and similarly it will add up in successive repeater sections for a long line. With present types of tubes operating with a moderate plate potential, the modulation requirement can be met only at relatively low output levels. To improve this situation and also to obtain advantages in amplifier stability, the reversed feedback principle employed for cable carrier amplifiers, as described in a paper by H. S. Black,[14] has been extended to higher frequency ranges. It has been found that amplifiers of this type having 30 db feedback reduce the distortion to such an extent that each amplifier of a long system carrying several hundred telephone channels will handle satisfactorily a channel output signal level about 5 db above that at the input of the toll line.

The maximum gain which can be used in the repeater, therefore, is, in the illustrative case given above of a long system carrying several hundred telephone channels, the difference between the minimum and maximum levels of 55 db below and 5 db above the point of reference, respectively, or a total gain of 60 db. (With a .3-inch coaxial line of the type shown in Fig. 2, this corresponds to a repeater spacing of about 10 miles.) If a repeater is to have 60 db net gain and at the

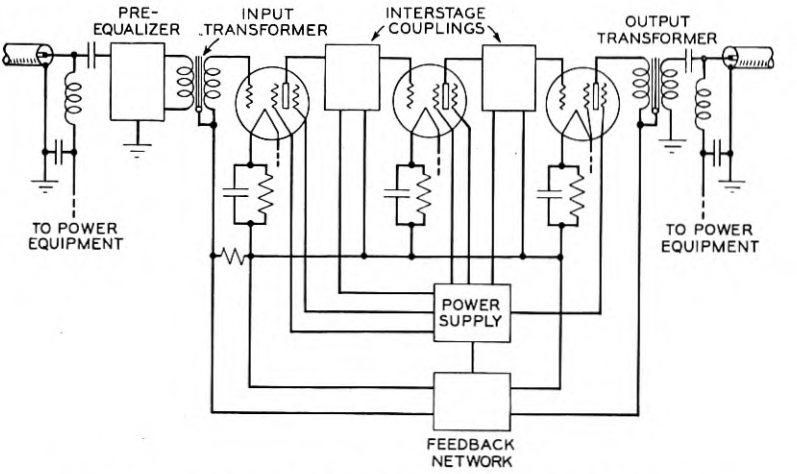


Fig. 10—Circuit of 1000-kilocycle three-stage feedback repeater.

same time about 30 db feedback, it is obvious that the total forward gain through the amplifying stages must be about 90 db. The circuit of an experimental amplifier meeting the gain requirements for a frequency band from 50 to 1000 kilocycles is shown schematically in Fig. 10.

Gain-Frequency Characteristic

As pointed out above, the line attenuation is not uniform with frequency. For a repeater section which has a loss of, say, 60 db at 1000 kilocycles, the loss at 50 kilocycles would be only about 15 db. Such a sloping characteristic can be taken care of either by designing the repeater to have an equivalent slope in its gain-frequency characteristic or by designing it for constant gain and supplementing it with an equalizer which gives the desired overall characteristic. Both methods have been tried out, as well as intermediate ones. Figure 11

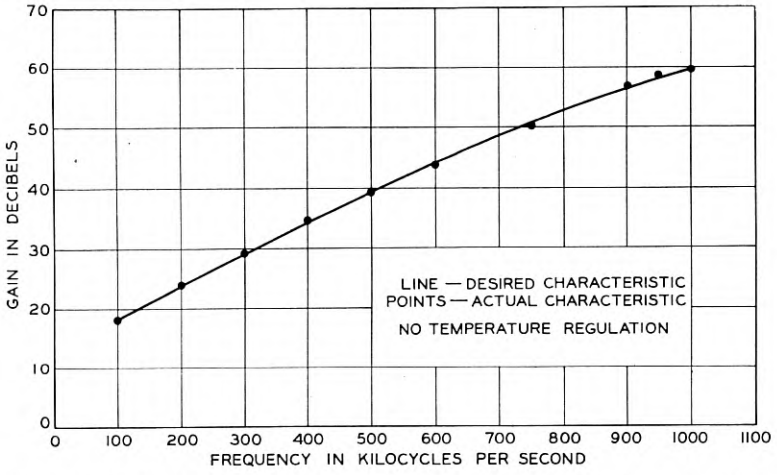


Fig. 11—Gain of 1000-kilocycle repeater compared with line characteristic.

illustrates such a sloping characteristic obtained by adjusting the coupling impedances in a three-tube repeater, designed in this case for 60 db gain at 1000 kilocycles. The accompanying photograph, Fig. 12, gives an idea of the apparatus required in such a repeater, apart from the power supply equipment.

Regulation for Temperature Changes

It is necessary that the repeater provide compensation for variations in the line attenuation due to changes of temperature. In the case of aerial construction such variations might amount to as much as 8 per cent in a day or 16 per cent in a year. If the line is under-

ground the annual variation is only about one-third of the above value and the changes occur much more slowly. On a transcontinental line the annual variation might total about 1500 db. Inasmuch as it is desirable to hold the transmission on a long circuit constant within about ± 2 db, it is obvious that the regulation problem is a serious one.

In a single repeater section of aerial line the variation might amount to ± 2.5 db per day or ± 5 db per year. Such variations, if allowed

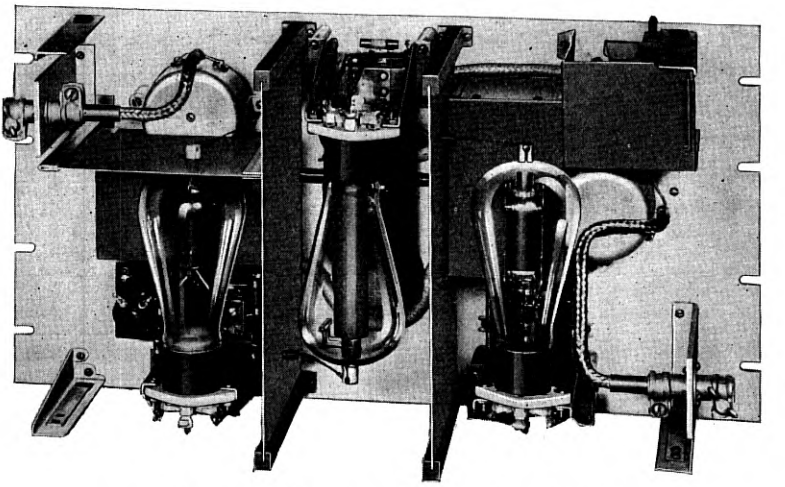


Fig. 12—Photograph of 1000-kilocycle repeater.

to accumulate over several repeater sections, will drop the signal down into the noise or raise it so as to overload the tubes. It is, therefore, advisable to provide some regulation at every repeater in an aerial line so as to maintain the transmission levels at approximately their correct position. For underground installations the regulating mechanism may be omitted on two out of every three repeaters.

In choosing a type of regulator system the necessity for avoiding cumulative errors in the large number of repeater sections has been borne in mind. In view of the wide band available, a pilot channel regulator system was naturally suggested. Such a scheme employing two pilot frequencies has been used experimentally to adjust the gain characteristic in such a way as to maintain the desired levels throughout the band. The accuracy with which this has been accomplished for a single repeater section is illustrated in Fig. 13. Over the entire band of frequencies and the extreme ranges in temperature which may be encountered, the desired regulation is obtained within a few tenths of a db.

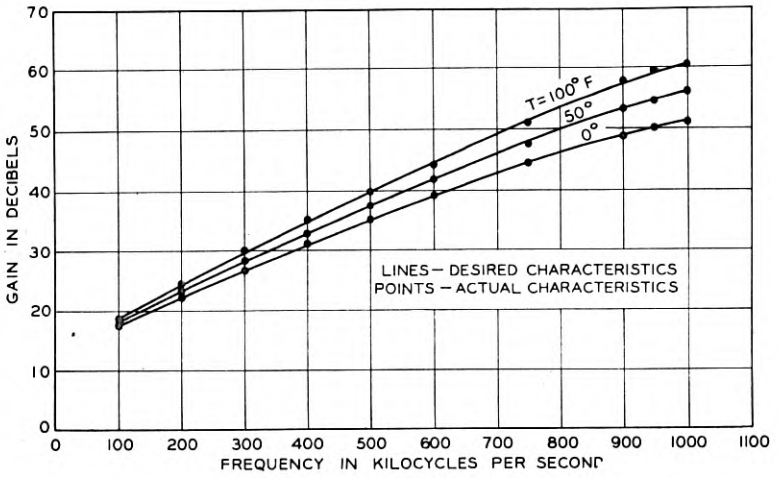


Fig. 13—Temperature regulation—line and repeater characteristics.

Repeater Operation, Power Supply, Housing, Etc.

In view of the large number of repeaters required in a broad-band transmission system it is essential that the repeater stations be simple and involve a minimum of maintenance. With the repeater design as described it is expected that most of the repeaters may be operated on an unattended basis, requiring maintenance visits at infrequent intervals.

An important factor in this connection is the possibility of supplying current to unattended repeaters over the transmission line itself. The coaxial line is well adapted to transmit 60-cycle current to repeaters without extreme losses and without hazard. The repeaters with regulating arrangements as built experimentally for a million-cycle system are designed to use 60-cycle current, which in this case appears to have the usual advantages over d.-c. supply. One repeater requires a supply of about 150 watts. The number of repeaters which can be supplied with current transmitted over the line from any one point depends upon the voltage limitation which may be imposed on the circuit from considerations of safety.

For a repeater of the type described with current supplied over the line, only a very modest housing arrangement will be required. For the great majority of stations, it appears possible to accommodate the repeaters in weatherproof containers mounted on poles, in small huts, or in manholes.

Higher Frequency Repeaters

Most of what has been said above applies particularly to repeaters transmitting frequencies up to about 1000 kilocycles. However, study has been given also to repeaters, both of the feedback and the non-feedback type, for transmitting higher frequencies. Experimental repeaters covering the range from 500 to 5000 kilocycles have been built and tested. These were capable of handling simultaneously the full complement of over 1000 channels which such a broad band will permit. The frequency characteristic of one of these repeaters, and the measured attenuation of a section of line of the type tested at Phoenixville are shown in Fig. 14.

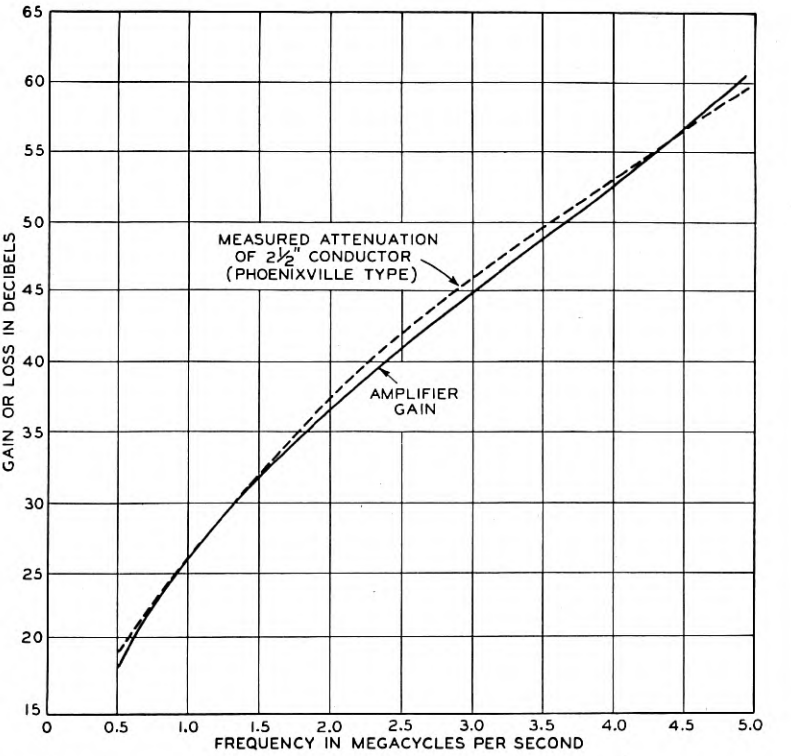


Fig. 14—Frequency characteristic of coaxial line and 5000-kilocycle repeater.

Terminal Arrangements

In order to utilize a broad band effectively for telephone purposes, the speech channels must be placed as close together in frequency as practicable. The factors which limit this spacing are: (1) The width of

speech band to be transmitted and (2), the sharpness of available selecting networks.

As to the width of speech band, the present requirement for commercial telephone circuits is an effective transmission band width of at least 2500 cycles, extending from 250 to 2750 cycles. It has been found that a band of this width or more may be obtained with channels spaced at 4000-cycle intervals. Band filters using ordinary electrical elements are available,[3] for selecting such channels in the range from zero to about 50 kilocycles. Channel selecting filters using quartz crystal elements [15, 16] have been developed in the range from about 30 to 500 kilocycles. The selectivity of a typical filter employing quartz crystal elements is shown on Fig. 15.

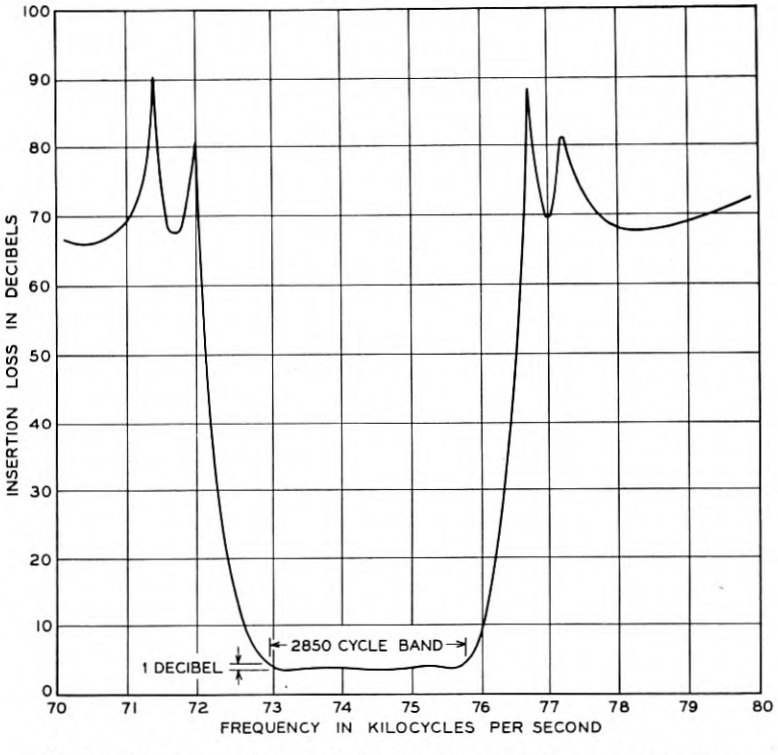


Fig. 15—Frequency characteristic of quartz crystal channel band filter.

Initial Step of Modulation

The initial modulation (from the voice range) may be carried out in an ordinary vacuum tube modulator or one of a number of other non-linear devices. The method chosen for the present experimental work

employs a single sideband with suppressed carrier, using a copper-oxide modulator associated with a quartz crystal channel filter. The terminal apparatus required for two-way transmission over a two-path circuit is shown diagrammatically on the left-hand side of Fig. 16.

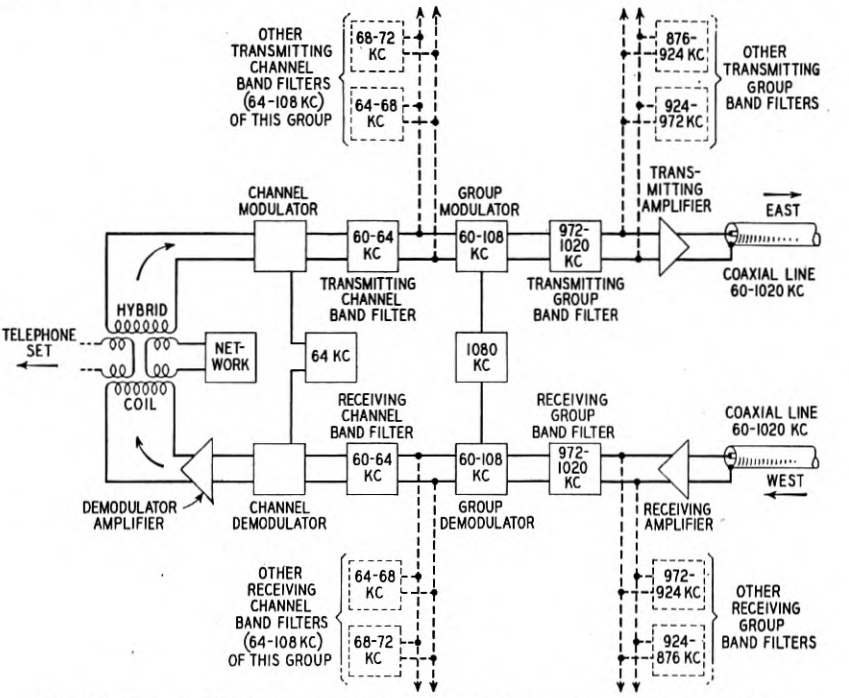


Fig. 16—Schematic of four-wire circuit employing two steps of modulation.

A frequency allocation which has been used for experimental purposes employs carriers from 64 to 108 kilocycles for the initial step of modulation. The lower sidebands are selected and placed side by side in the range from 60 to 108 kilocycles, as illustrated in Fig. 17, forming a group of 12 channels.

Double Modulation

In order to extend the frequency range of a system to accommodate a very large number of channels, it appears to be more economical to add a second step of modulation rather than carry the individual channel modulation up to higher frequencies. Such a second step of modulation has been used experimentally to translate the initial group of 12 channels en bloc from the range 60 to 108 kilocycles up to higher frequencies. It is possible to place such groups of channels one above another as illustrated in the upper part of the diagram of Fig. 18, up

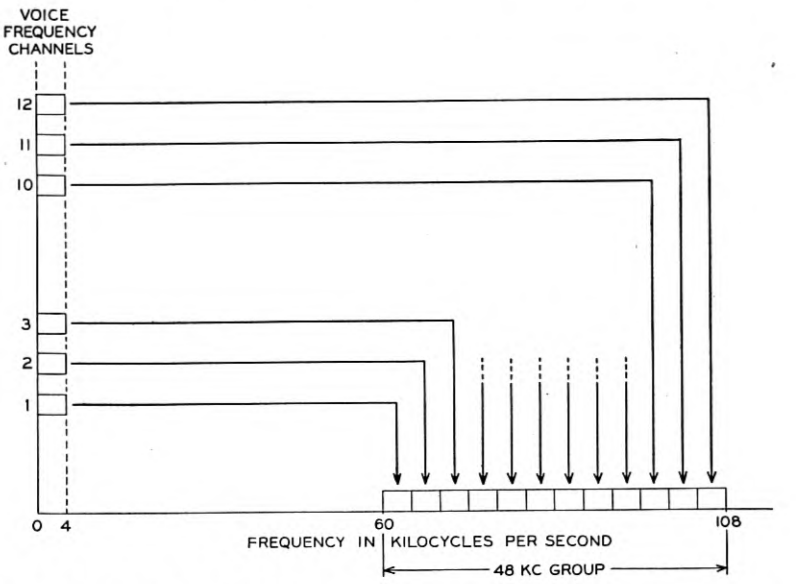


Fig. 17—Diagram illustrating frequency allocation for first step of modulation.

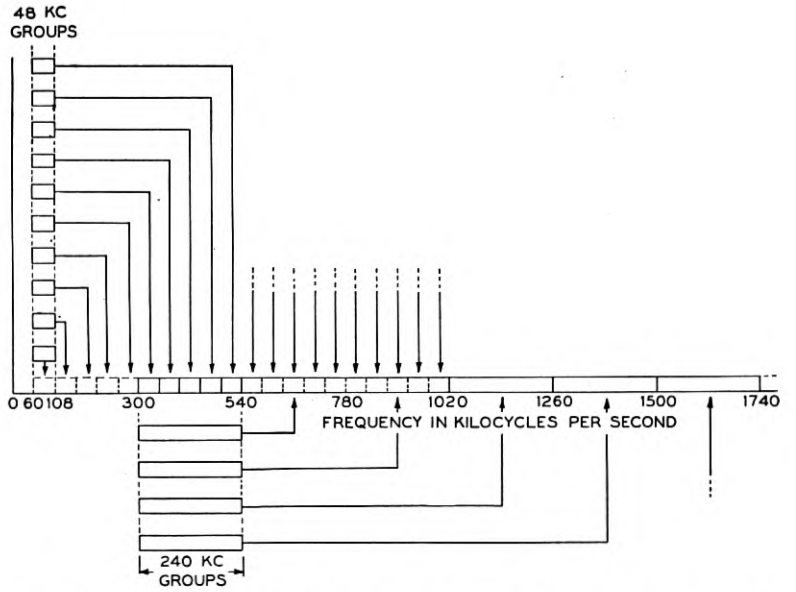


Fig. 18—Diagram illustrating frequency allocation for two or three steps of modulation.

to about 1000 kilocycles, wasting no frequency space between groups and thus keeping the channels spaced at intervals of 4 kilocycles throughout the entire range.

The apparatus required for this purpose is shown schematically in Fig. 16, which illustrates the complete terminal arrangements for a single channel employing double modulation. The figure indicates by dottled lines where the other channels and groups of channels are connected to the system.

A modulator for shifting the frequency position of a group of channels inherently yields many different modulation products as a result of the intermodulation of the signal frequencies with the carrier frequency and/or with one another. Out of these products only the "group sideband" is desired. The number of the modulation products resulting merely from the lower ordered terms of the modulator response characteristic is extremely large. All such products must be considered from the standpoint of interference either with the group which is wanted in the output or with other groups to be transmitted over the system. Various expedients may be used to avoid interference as follows: (1) A proper choice of frequency allocation will place the undesired modulation products in the least objectionable location with respect to the wanted signal bands; (2) a high ratio of carrier to signal will minimize all products involving only the signal frequencies; (3) the use of a balanced modulator will materially reduce all products involving the second order of the signal; (4) selectivity in the group filters will tend to eliminate all products removed some distance from the wanted signal group. Giving due regard to these factors, balanced vacuum tube group modulators have been developed which are satisfactory for the frequency allocations employed.

Triple Modulation

For systems involving frequencies higher than about 1000 kilocycles it may be desirable to introduce a third step of modulation. In some experiments along this line a "super-group" of 60 channels, or five 12-channel groups, has been chosen. The lower part of Fig. 18 illustrates, for a triple modulation system, the shifting of supergroups of 60 channels each to the line frequency position. This method has been employed experimentally up to about 5,000 kilocycles. It is of interest to note that even in extending these systems to such high frequencies, channels are placed side by side at intervals of 4000 cycles to form a practically continuous useful band for transmission over the line.

Demodulation

On the receiving side the modulation process is reversed. The apparatus units are similar to those used on the transmitting side, and are similarly arranged. Figure 16 illustrates this for the case of double modulation.

Carrier Frequency Supply

In systems operating at higher frequencies it is necessary that the carrier frequencies be maintained within a few cycles of their theoretical position in order to avoid beat tones or distortion of the speech band. Separate oscillators of high stability could, of course, be used for the carrier supply but it appears more economical to provide carriers by means of harmonic generation from a fundamental basic frequency. Such a base frequency may be transmitted from one end of the circuit to the other, or may be supplied separately at each end.

Television

The broad band made available by the line and repeaters may be used for the transmission of signals for high-quality television. Such signals may contain frequency components extending over the entire range from zero or a very low frequency up to a million or more cycles.[4] The amplifying and transmitting of these frequencies, particularly the lower ones, presents a serious problem. The difficulty can be overcome by translating the entire band upward in frequency to a range which can be satisfactorily transmitted. To effect such a shift, the television band may first be modulated up to a position considerably higher than its highest frequency and then with a second step of modulation be stepped down to the position desired for line transmission.

This method is illustrated in Fig. 19 for a 500-kc. television signal band. The original television signal is first modulated with a relatively high frequency, two million cycles in this case (C_1). The lower sideband, extending to 1500 kilocycles, is selected and is modulated again with a frequency of 2100 kilocycles (C_2). The lower sideband of 100 to 600 kilocycles is selected with a special filter so designed that the low frequency end is accurately reproduced. The television signal then occupies the frequency range of 100 to 600 kilocycles as shown on the diagram and may be transmitted over a coaxial or other high frequency line. At the receiving end a reverse process is employed. The same method using correspondingly higher frequencies may be used for wider bands of television signals.

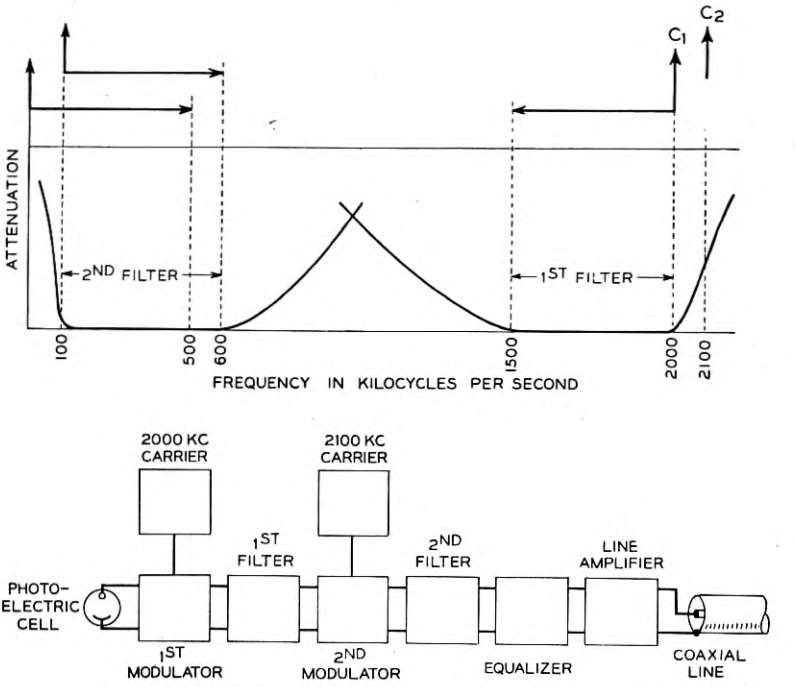


Fig. 19—Double modulation method for translating television signals for wire line transmission.

Other Communication Facilities

The telephone channels provided by the system may be used for other types of communication services, such as multi-channel telegraph, teletype, picture transmission, etc. For the transmission of a high-quality musical program, which requires a wider band than does commercial telephony, two or more adjacent telephone channels may be merged. The adaptability of the broad-band system to different types of transmission thus will be evident.

As already noted, the commercial application of these systems for wide-band transmission over coaxial lines must await a demand for large groups of communication facilities or for television. The results which have been outlined are based upon development work in the laboratory and the field, and it is probable that the systems when used commercially will differ considerably from the arrangements described.

REFERENCES

1. E. H. Colpitts and O. B. Blackwell, "Carrier Current Telephone and Telegraphy," *A. I. E. E. Trans.*, Vol. 40, February 1921, p. 205–300.
2. H. A. Affel, C. S. Demarest, and C. W. Green, "Carrier Systems on Long Distance Telephone Lines," *A. I. E. E. Trans.*, Vol. 47, October 1928, 1360–1367. *Bell Sys. Tech. Jour.*, Vol. VII, July 1928, p. 564–629.
3. A. B. Clark and B. W. Kendall, "Communication by Carrier in Cable," *Elec. Engg.*, Vol. 52, July 1933, p. 477–481, *Bell Sys. Tech. Jour.*, Vol. XII, July 1933, p. 251–263.
4. P. Mertz and F. Gray, "Theory of Scanning and Its Relation to the Characteristics of the Transmitted Signal in Telephotography and Television," *Bell Sys. Tech. Jour.*, Vol. XIII, July 1934, p. 464.
5. E. W. Engstrom, "A Study of Television Image Characteristics," *Proc. I. R. E.*, Vol. 21, December 1933, p. 1631–1651.
6. S. A. Schelkunoff, "The Electromagnetic Theory of Coaxial Transmission Lines and Cylindrical Shields." *Bell Sys. Tech. Jour.*, Vol. XIII, October 1934.
7. J. R. Carson and J. J. Gilbert, "Transmission Characteristics of the Submarine Cable," *Jour. Franklin Institute*, Vol. 192, December 1921, p. 705–735.
8. W. H. Martin, G. A. Anderegg and B. W. Kendall, "The Key West-Havana Submarine Telephone Cable System," *Trans. A. I. E. E.*, Vol. 41, 1922, p. 1–19.
9. British Patent No. 284,005, C. S. Franklin, January 17, 1928.
10. E. J. Sterba and C. B. Feldman, "Transmission Lines for Short-Wave Radio Systems," *I. R. E. Proc., Vol.* 20, July 1932, p. 1163–1202; also *Bell Sys. Tech. Jour.*, Vol. II, July 1932, p. 411–450.
11. U. S. Patents No. 1,835,031, L. Espenschied and H. A. Affel, December 9, 1931, and No. 1,941,116, M. E. Strieby, December 26, 1933.
12. E. I. Green, "Transmission Characteristics of Open-Wire Lines at Carrier Frequencies," *A. I. E. E. Trans.*, Vol. 49, October 1930, p. 1524–1535; *Bell Sys. Tech. Jour.*, Vol. IX, October 1930, p. 730–759.
13. H. A. Affel and E. I. Green, U. S. Patent No. 1,818,027, Aug. 11, 1931.
14. H. S. Black, "Stabilized Feed-Back Amplifiers," Electrical Engineering, Vol. 53, January 1934, p. 114–120. *Bell Sys. Tech. Jour.*, Vol. XIII, January 1934, p. 1–18.
15. L. Espenschield, U. S. Patent No. 1,795,204, March 3, 1931.
16. W. P. Mason, "Electrical Wave Filters Employing Quartz Crystals as Elements," *Bell Sys. Tech. Jour.*, Vol. XIII, July 1934, p. 405.

Regeneration Theory and Experiment *

By

E. PETERSON, J. G. KREER, AND L. A. WARE

A comprehensive criterion for the stability of linear feed-back circuits has recently been formulated by H. Nyquist, in terms of the transfer factor around the feed-back loop. The importance of any such general criterion lends interest to an experimental verification, with which the paper is primarily concerned.

The subject is dealt with under five principal headings. The first section reviews some of the criteria for oscillation to be found in the literature of vacuum tube oscillators. The second describes the derivation of Nyquist's criterion somewhat along the lines followed by Routh in one of his investigations of the stability of dynamical systems. The third part deals with two experimental methods used in measuring the transfer factor. The fourth is concerned with the particular amplifier circuit used in the test of Nyquist's criterion. The last section applies the criterion to a nonlinear case, and to circuits including two-terminal negative impedance elements.

IN a comparatively recent paper on "Regeneration Theory," [1] Dr. Nyquist presented a mathematical investigation of the conditions under which instability [2] exists in a system made up of a linear amplifier and a transmission path connected between its input and output circuits. The results of the investigation are of interest because of their obvious application to amplifiers provided with feed-back paths,[3] as well as to the starting conditions in oscillators. As a result of his general analysis, Dr. Nyquist arrived at a criterion for stability, expressed in particularly simple and convenient form, which is not restricted in its range of application to particular amplifier and circuit configurations.

The great value attached to a criterion as precise and as general as Nyquist's makes it desirable to submit the criterion to an experimental test. One particularly striking conclusion drawn from this criterion is that under certain conditions a feed-back amplifier may sing within certain limits of gain, but either reduction or increase of gain beyond these limits may stop singing. A feed-back amplifier satisfying these conditions was set up, and the experimental results were found to be in agreement with this conclusion.

* Published in *Proc. I. R. E.*, October, 1934.

[1] *Bell. Sys. Tech. Jour.*, vol. XI, p. 126.

[2] Instability is used in the sense that a small impressed force, which dies out in course of time, gives rise to a response which does not die out.

[3] *Electrical Engineering*, July, 1933; *Bell Sys. Tech. Jour.*, p. 258, July, 1933.

It is interesting to compare the criterion with those derived for the mechanical systems of classical dynamics. In his Adams Prize Paper on "The Stability of Motion," [4] and again in his "Advanced Rigid Dynamics," [5] Routh investigated the general problem of dynamic stability and established a number of criteria based upon various properties of dynamical systems. When applied to the problem of feed-back amplifiers, keeping Nyquist's result in mind, one of them is found to be equivalent to Nyquist's criterion, although expressed in different terms and derived in a different way.

To provide a background for the experiments, we propose to state some of the criteria for stability which are to be found in the literature of vacuum tube oscillators, and to compare them with Nyquist's or Routh's criterion, the development of which is most conveniently described somewhat along the lines followed by Routh. Following this we shall deal with the experimental methods and apparatus which were used in testing the criterion, and conclude with some extensions of the criterion.

Circuit Analysis and Stability

Conditions required for the starting of oscillations in linear feedback circuits, corresponding to instability, are to be found in the literature of vacuum tube oscillator circuits, expressed in a number of ostensibly different forms. These are usually based upon the familiar mesh differential equations for the system which involve differentiations and integrations of the mesh amplitudes with respect to time. Using the symbol p to denote differentiation with respect to time, each mesh equation becomes formally an algebraic one in p, involving the circuit constants and the mesh amplitudes. The solution of this system of equations is known to be expressible as the sum of steady state and transient terms. The transient terms are each of the form $B_k \epsilon^{p_k t}$, the B_k's being fixed by initial conditions, and the p_k's being determined from the circuit equations. If we set up the determinant of the system of equations—the discriminant—and equate it to zero, the roots of the resulting equation are the p_k's above. In general each mesh equation involves p to the second degree at most, and with n meshes the discriminant is of degree $2n$ at most. Accordingly we may express the determinantal equation as

$$F(p) = 0 = K(p - p_1)(p - p_2) \cdots (p - p_{2n}). \tag{1}$$

As for the steady state term, in the simplest case in which a sinusoidal wave of frequency $\omega/2\pi$ is impressed, it is equal to the impressed

[4] Macmillan, 1877.
[5] Macmillan, 6th edition, 1905.

voltage divided by the discriminant and multiplied by the appropriate minor of the determinant, in which p is replaced by $j\omega$. The character of the response due to a slight disturbance and in the absence of any periodic force is determined by the exponentials. In general, p_k is a complex quantity which may be written as $a_k + j\omega_k$. It is apparent that in the critical case for which a_k is zero, the corresponding term becomes $\epsilon^{j\omega_k t}$, corresponding to an oscillation invariable in amplitude, of frequency $\omega_k/2\pi$. If a_k is negative, as is ordinarily the case when the system is passive (containing no amplifier or negative impedance), then the oscillation diminishes in course of time. When a_k is positive, however, the oscillation increases with time, and the system is said to be unstable. Evidently the stability of a system is determined by the signs of the a_k's.

Several criteria which have previously been enunciated for the maintenance of free oscillations are deducible from the above. One states that the discriminant must vanish when p takes on the value $j\omega$. Another states that the damping (a_k) must be zero at the frequency of oscillation. These are clearly equivalent. Two derived criteria may also be mentioned, based upon the properties of the system when the circuit is broken. The first of these states that if the impedance is measured looking into the two terminals provided by the break, the impedance must be zero at the frequency of steady oscillation.

The second criterion involving the transfer factor has become fairly widespread, perhaps because it leads to a simple and plausible physical picture. To determine the transfer factor around the feed-back loop, the loop is broken at a convenient point, and the two sets of terminals formed by the break are each terminated in a passive impedance equal to that which is connected in the normal (unbroken) condition. Then when a voltage of frequency $\omega/2\pi$ is applied to one of the pairs of terminals so provided—the input terminals [6]—and the corresponding voltage is measured across the other pair, the transfer factor $A(j\omega)$ is obtained as the vector ratio of the output voltage to the input voltage.

The manner in which the transfer factor enters into the problem may be demonstrated directly by comparing the voltages at any point of the main amplifier circuit under the two conditions in which the feed-back path is opened and closed respectively. If with the feed-back path open the voltage at any such point is $E\epsilon^{pt}$, then when the feed-back path is closed the voltage will be changed [7] to

$$E\epsilon^{pt}/[1 - A(p)].$$

[6] Input terminals are those across which an impressed potential leads to propagation in the normal direction of amplifier transmission.

[7] *Bell Sys. Tech. Jour.*, Vol. XI, p. 128.

This may be shown as follows with reference to the particular circuit of Fig. 1: If the feed-back circuit is broken and then properly terminated, the voltage existing across the input is taken as e. Now suppose the feed-back path to be restored. Designating the voltage existing across the input in the presence of feed-back as e_1 we have $e_1 = e + Ae_1$, from which the above equation follows.

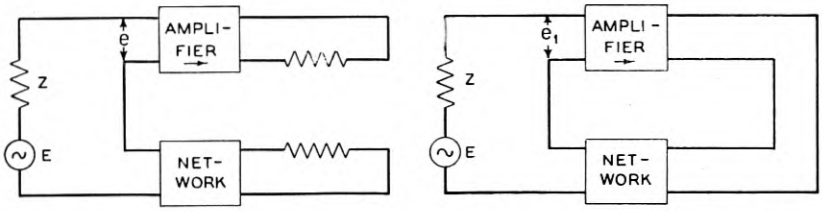


Fig. 1—Series type feed-back; loop broken and terminated at left, normal feed-back circuit at right.

If we let $F_1(p)$ represent the discriminant of the system when the loop is broken and terminated, then the roots of the equation formed by setting the discriminant equal to zero are assumed to have positive real parts. Now for the corresponding discriminant when the loop is restored, we have in accordance with the above considerations

$$F(p) = [(1 - A(p)]F_1(p).$$

In setting this discriminant equal to zero to obtain the roots, the only ones which have nonnegative real parts are those corresponding to the feed-back term

$$f(p) = 1 - A(p). \tag{2}$$

The above-mentioned criterion may be deduced from this expression. For steady oscillations to exist the output potential must be identical in amplitude and in phase with that existing across the input at the frequency of oscillation ($p = j\omega$), in which case the transfer factor is unity. This seems reasonable on the basis that when the input and output terminals are connected through, the oscillation will neither increase nor decrease with time. It may be demonstrated by direct analysis that these several criteria, framed for the critical case of undamped oscillations, all lead to the same correct conclusion.

Of course in any actual oscillating circuit it is practically impossible to get these conditions fulfilled exactly, and what is ordinarily done in the practical design of oscillating circuits is to ensure that the voltage fed back will be greater than that required to produce oscillation. This evidently goes a step further than the above criteria, and

reliance is placed upon the nonlinear properties of the circuit to fulfill the criteria automatically. The procedure is known by experiment to be effective in the usual type of oscillating circuit. In particular forms of feed-back circuits, however, it may be demonstrated that *the transfer factor may be made greater than unity without giving rise to oscillations*. This situation was investigated experimentally, and found to be in accord with the stability criterion stated by Nyquist.

Nyquist's Criterion

The explicit solution of (1) for the p_k's demands an exact knowledge of the configuration of the amplifier and feed-back circuits. When the number of meshes is large, the solution involves much labor. If we wish simply to observe whether or not the system is stable, however, we need not obtain explicit solutions for the roots; in fact, all we need to know is whether or not any one of the p_k's has its real part positive. It turns out that when we know the transfer factor as a function of frequency, by calculation or by measurement, a simple inspection of the transfer factor polar diagram suffices for this purpose. This diagram is constructed by plotting the imaginary part of the transfer factor against the real part for all frequencies from minus to plus infinity.[8]

To obtain Nyquist's criterion we consider the vector drawn from the point (1, 0) to a point moving along the polar diagram; if the net angle which the vector swings through in traversing the curve is zero, the system is stable; if not, it is unstable. To express it in the terms used by Routh, if we set $1 - A(j\omega) = P + jQ$, and observe the changes of sign which the ratio P/Q makes when P goes through zero as the frequency steadily increases, the system is stable when there are the same number of changes from plus to minus as from minus to plus. It may be demonstrated that these two statements are equivalent.

The way in which the above procedures may be shown to reveal the existence of a root with positive real part may be outlined somewhat along the lines followed by Routh in his analysis.[9] Since p is a com-

[8] The transfer factor for negative frequencies $A(-j\omega)$ is the complex conjugate of that for positive frequencies $A(j\omega)$. Thus, if

$$A(j\omega) = X + jY,$$

then

$$A(-j\omega) = X - jY.$$

[9] A number of restrictions on the generality of the analysis may be noted. It is assumed that $A(p)$ has no purely imaginary roots, although the result in this case is otherwise evident. Further it is assumed that $A(p)$ goes to zero as $|p|$ becomes infinite, and that no negative resistance elements are included in the amplifier. Another point which should be mentioned is that the analysis does not apply to the stability in any conjugate paths that may exist. This point may be exemplified by the balanced tube or push-pull amplifier, in the normal transmission path of which the tubes of a stage act in series. When the series output is connected back to the series

plex quantity in general, any value which it may take is representable as a point on a plane—the p-plane of Fig. 2. Since only values of p with positive real parts concern us, attention may be confined to the right-hand half of the p-plane. Now draw a closed contour C in the right-hand half of the p-plane which encloses the root p_k. It is evident

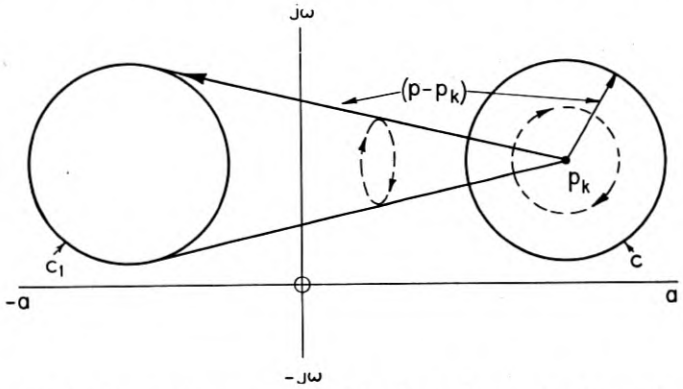


Fig. 2—Plot of two contours C and C_1 in the p-plane. C encloses the root p_k while C_1 does not enclose p_k. The vector $p - p_k$ covers 360 degrees as p traverses C, and covers the net angle of zero as p traverses C_1.

upon inspection of Fig. 2 that the vector extending from the root p_k to the contour makes a complete revolution (360 degrees) in following the closed path. If the contour does not enclose the root, however, as for C_1, then it is clear that when the vector from the root to the contour traverses the whole contour, the net angle turned through by the vector is zero. In the region under consideration we may write

$$f(p) = (p - p_k)\phi(p),$$

where $\phi(p)$ has no zeros within the contour. Hence, when p traverses a closed path and $(p - p_k)$ turns through 360 degrees or through zero the same angle is covered by $f(p)$. If for some different contour several roots are enclosed, it may be shown that $f(p)$ turns through one complete revolution for each of the enclosed roots when p traverses the contour.

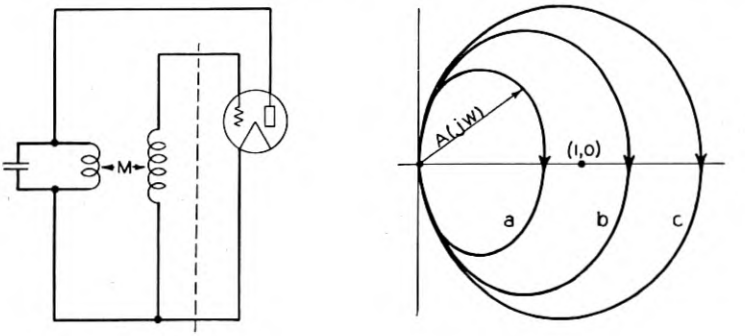
In the form in which these considerations are stated, they are not suitable to practical application since complex values of p are involved. Ordinarily, of course, only imaginary values ($p = j\omega$) are conveniently accessible to us since it is a comparatively simple matter to

input, stability of the resultant loop has in general no bearing upon the stability of the path formed with the two tubes of each stage in parallel, since the series and shunt paths are conjugate to one another. To establish the stability of the shunt or parallel path, the transfer factor for that path must be separately determined. In general, the stability criterion applies only to the particular loop investigated, and not to any other existent loop.

measure the response with a sinusoidal impressed wave, but it would involve great difficulties of experiment as well as of interpretation to determine the response with negatively damped waves corresponding to values of p in the right-hand half of the plane. However, these results may be brought within the field of practical experience by a procedure widely used for the purpose.

To include all roots in the right-hand half of the p-plane, the contour must be taken of infinite extent. The path ordinarily followed for this purpose extends from the value $+ R$ to $- R$ on the imaginary axis, and is closed by a semicircle of radius R, where R is assumed to expand without limit. It may be noted that in actual amplifier circuits the transfer factor becomes zero when $|p|$ becomes infinite, so that $A(p)$ is zero along the semicircular part of the closed contour. Consequently, the only values of $A(p)$ which differ from zero are those corresponding to finite values of p, along the imaginary axis. In other words, the plot of $A(p)$ under these conditions comes down to the plot of $A(j\omega)$ where ω is finite. Hence, if we plot $A(j\omega)$ for all values of ω from minus to plus infinity, there will be no roots with positive real parts and the system will be stable when the vector from (1, 0) to the curve sweeps through a net angle of zero. The system will be unstable when the vector sweeps through 360 degrees, or an integral multiple thereof.

Two types of transfer factor curves may be considered as illustrations. The first of these shown in Fig. 3 corresponds to that for a re-



a b

Fig. 3—Schematic of a reversed feed-back oscillator circuit at the left. At the right plot of the transfer factor $A(j\omega)$ around the feed-back loop of Fig. 3a over the frequency range from zero to very high frequencies. The imaginary part of the transfer factor is plotted as ordinate against the real part as abscissa for the three curves *a*, *b*, *c*, which correspond to increasing gains around the loop. Condition *a* is stable, while *b* and *c* are unstable.

versed feed-back oscillator circuit, the three curves marked *a*, *b*, *c*, corresponding to progressively increasing gains around the loop. It will be observed that after the maximum gain has reached and exceeded unity, that the circuit is unstable, since the point (1, 0) is then enclosed. This state of affairs may be contrasted with that existing in the particular form of feed-back circuit to which Fig. 4 applies. Again the

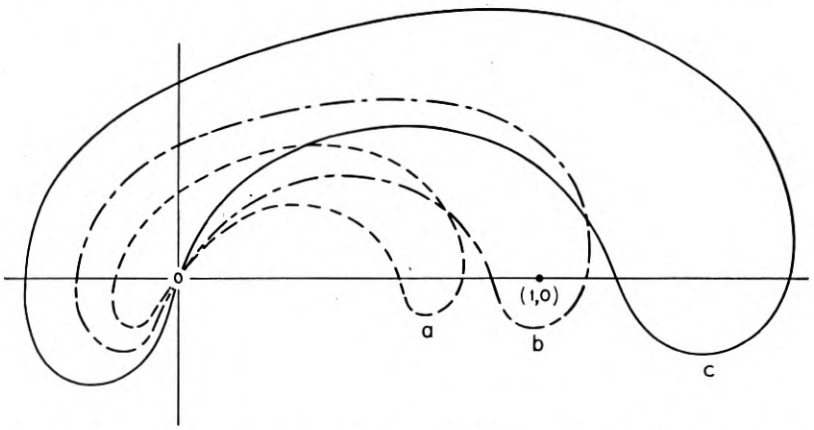


Fig. 4—Transfer factor diagram for a particular form of feed-back circuit, curves *a*, *b*, *c*, corresponding to increasing gains around the feed-back loop. Conditions *a* and *c* are stable, *b* is unstable.

three curves *a*, *b*, *c*, correspond to progressively increasing gains around the feed-back loop. As the gain is increased the system is first stable (*a*), then unstable (*b*), and finally stable (*c*), since it is only within curve (*b*) that the point (1, 0) is enclosed. This striking example is the one which was investigated experimentally. The methods used in determining the transfer factor diagram form the subject of the next section.

Measuring Methods

Application of the Nyquist stability criterion requires the determination of the vector transfer factor around the feed-back loop at all frequencies. This is usually effected by opening the circuit at any point which provides convenient impedances looking in both directions from the break. These points are then connected to an oscillator and to suitable measuring circuits, which are to be described. Care must be taken to ensure that the oscillator and measuring circuit impedances are equal to the output and input impedances respectively of the circuit under test. This precaution is necessary in order that the transfer factor in the measuring condition may not differ significantly from that existing in the operating condition.

Two methods of measurement have been found useful. The first is a null method capable of good precision over a wide frequency range. The second is a visual method in which the transfer factor polar diagram is traced on the screen of a cathode ray oscillograph. This method is not capable of very great precision and, in the model used, the frequency range is somewhat restricted. However, it permits of a rapid survey of the situation for which its precision is adequate, before proceeding with the slower and more precise measurements of the null method, where the latter are required. By making such a preliminary survey the critical frequency ranges can be mapped out for precise measurement, thereby eliminating a large amount of unnecessary labor.

Null Method

In the more precise measurements extending over a wide frequency range, special care is required to ensure freedom from errors in the measurement of phase angles and amplitudes. Much of the difficulty associated with direct measurement over wide frequency ranges is avoided by the use of a simple demodulation scheme. In this scheme, the potentials to be compared are modulated down to a fixed frequency (in actual use 1000 cycles) regardless of the frequency at which the test is being made. In this way a minimum portion of the circuit carries the high frequency. Further this permits the use of voice frequency attenuators, phase shifters, and amplifiers which in fact require calibration at only a single frequency.

In this arrangement, as shown in Fig. 5, demodulators are shunted across the input and output terminals of the circuit under test. A single oscillator supplies the carrier to both demodulators, its frequency differing by 1000 cycles from the frequency supplied to the circuit under test. The demodulated outputs are connected through attenuators and phase shifters to a common amplifier detector. The attenuators and phase shifters are adjusted until the detector gives a null reading. When this condition obtains the difference in the attenuator settings in the two branches is equal to the gain or loss of the circuit under test, and the difference in the phase shifter settings is either equal to or the negative of the phase shift of the circuit under test. To show this, denote the amplifier output voltage by $P_0 \cos (2\pi ft - \phi)$, and the beat frequency voltage supplied to the demodulators by $P \cos 2\pi(f \pm 1000)t$. The demodulated output, proportional to the product of the two applied waves, is then

$$PP_0 \cos (2\pi \cdot 1000t \mp \phi).$$

Correspondingly, the demodulated output from the other demodulator

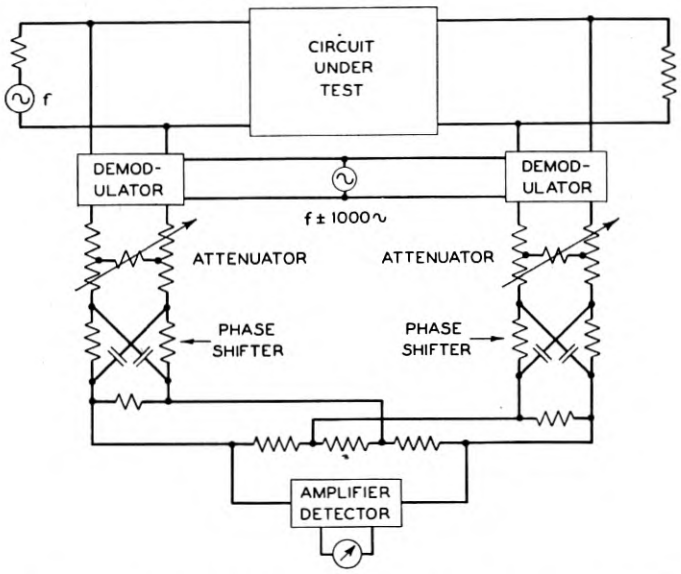


Fig. 5—Schematic diagram of the null method used to measure the transfer factor.

connected across the input is given by

$$PP_i \cos (2\pi \cdot 1000t).$$

If now these two waves are to be made to cancel, there must be a difference in the attenuation of the two branches equal to the ratio P_0/P_i, and a difference in the phase shift equal to $\mp \phi$. The change in sign of the phase angle introduced by setting the beat oscillator above or below the test frequency is most conveniently handled by setting the carrier oscillator consistently on the same side of the test frequency in making a run over the frequency range.

By using a high gain amplifier preceding the detector, the precision may be made great, limited only by circuit noise and by interference. The attenuators and phase shifters are calibrated separately. It should be noted that any difference in the transfer constants of the two demodulator circuits may be compensated by an initial adjustment which is carried out by paralleling the input terminals of the two demodulators across a source of electromotive force. With the particular type of phase shifter used the phase shift may be changed without altering the attenuation, so that the two settings for amplitude and phase may be made independently.

Visual Method

In the visual method of observation, a steady potential proportional to the inphase component of the transfer factor is impressed across one pair of plates of a cathode ray oscillograph and another steady potential proportional to the quadrature component is impressed across the other pair of plates, the constant of proportionality being the same for the two components. In this way the transfer factor at any frequency appears as a single point, the vector from the origin to the displaced beam constituting the transfer factor. The locus of all these points, i.e., vector tips, over the frequency range constitutes the transfer factor polar diagram.

To provide rectified potentials proportional to inphase and to quadrature components respectively, use is made of the properties of the so-called vacuum tube wattmeter.[10] As used in practice, this device consists of two triodes in push-pull connection (Fig. 6), the series arm of the grid circuit being connected to the unknown potential, and the

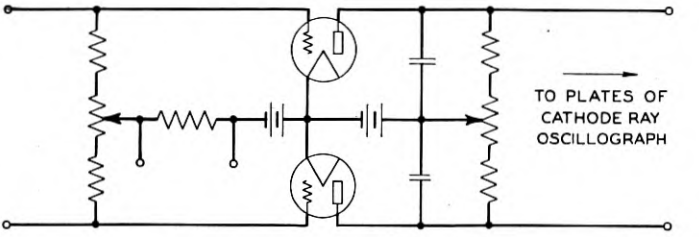


Fig. 6—Circuit of a vacuum tube wattmeter used to provide a rectified potential proportional to the product of the two impressed grid potentials (both of the same frequency) multiplied by the cosine of the phase angle between them.

shunt arm of the grid circuit being connected to a source of the same frequency but of standard phase. Under these conditions the rectified output in the plate circuit flowing in series with the two plates is proportional to the product of the two impressed voltages multipled by the cosine of the angle between them.

As shown in Fig. 7, two separate wattmeters are employed, one for each phase, their series input terminals being connected together across the output of the circuit under test. To the common branch of one of these wattmeters is supplied the same potential as is fed to the input of the circuit under test. The rectified output of this wattmeter therefore is proportional to the product of the input and output voltages multiplied by the cosine of the transfer factor phase angle. This po-

[10] U. S. Patent 1,586,533; Turner and McNamara, *Proc. I. R. E.*, vol. 18, p. 1743; October (1930).

tential is supplied to those plates of the oscillograph which produce a horizontal deflection. To the common branch of the other wattmeter is applied a potential equal in amplitude to the input voltage but lagging behind it by 90 degrees. The rectified output of this wattmeter is proportional to the product of input and output voltages multiplied by the cosine of the transfer factor phase angle minus 90 degrees, or in other words proportional to the sine of the transfer factor phase angle. This voltage is supplied to those plates of the oscillograph which produce a vertical deflection. We have then across one pair of plates of the oscillograph a steady potential proportional to the real component

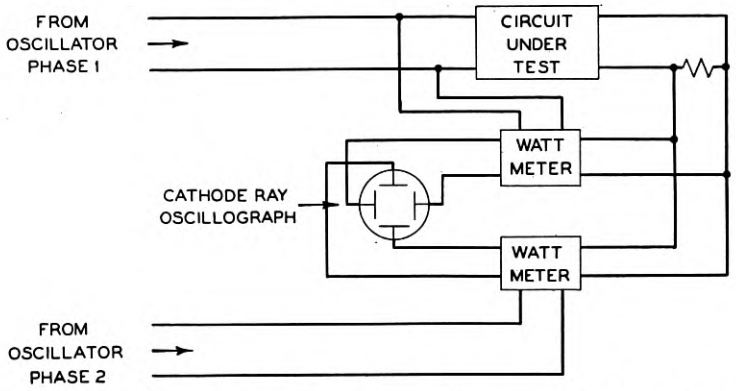


Fig. 7—Schematic diagram of the circuit used to plot the transfer factor diagram on the screen of a cathode ray oscillograph.

of the transfer factor, and across the other pair of plates we have impressed a steady potential proportional to the imaginary component of the transfer factor. These two components act upon the beam of the oscillograph to produce a deflection which in amplitude and in phase is the resultant of the two component deflections and so corresponds to the transfer factor.

It will be observed that the above procedure requires a two-phase source of constant amplitude, the frequency of which is variable over the range necessary to establish the properties of the amplifier. In the present instance the frequency range extends from 0.5 to 30 kilocycles, and the accuracy required is of the order of five per cent.

A schematic of the two-phase oscillator used is shown in Fig. 8. This oscillator is of the heterodyne type. Two independent sources are used, one of constant frequency (100 kilocycles), the other variable in frequency and practically constant in amplitude over the range of 100 to 130 kilocycles. As indicated in the figure, the variable fre-

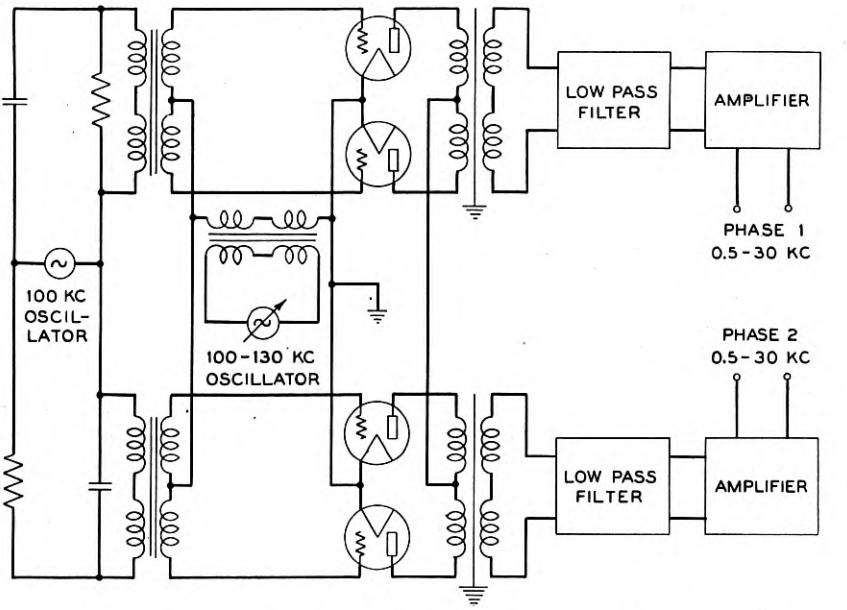


Fig. 8—Circuit diagram of a heterodyne type two-phase oscillator, the output frequency of which is continuously variable from 0.5 to 30 kilocycles. The output of each phase and the 90-degree difference between the two phases are practically constant over the frequency range.

quency oscillator is connected to the common branches of the two push-pull modulators. The fixed frequency oscillator is connected in series with the grid circuits of the two modulators. The resistance-capacity networks shown in the circuits of the fixed frequency oscillator are provided to produce phase shifts of 90 degrees between the two series voltages of the two modulators. In the same manner as that discussed before in connection with the null method measuring circuit, the phase shift introduced to the fixed frequency is maintained in the beat frequency output, so that the phase difference of 90 degrees is preserved in the outputs of the two modulators when the variable frequency oscillator goes from about 100.5 to 130 kilocycles. The outputs of the two phases are connected to the test amplifier and to the wattmeters as shown in the preceding Fig. 7.

Comparison of the Methods

Measurements of transfer factors by the two methods outlined above were found to be in agreement within the error of measurement. The visual method as developed was capable of use over only a very re-

stricted frequency range as compared to the null method, but it covered the region of particular interest in the experiments conducted for the purpose of testing the stability criterion. Through its use, measurements over its frequency range could be made in a few minutes time, whereas corresponding measurements by the more precise null method required three to six hours. Of course the time intervals cited do not include time occupied in setting up and adjusting the apparatus.

Test Amplifier and Experimental Results

Test Amplifier

The stability criterion indicates three distinct conditions of interest, one of which is unstable, the other two being stable. The unstable condition (1) is that in which the transfer factor curve encloses the point (1, 0). Two stable conditions are those in which (1, 0) is not enclosed by the curve, but in which (2) the curve crosses the zero phase shift axis at points greater than unity, and (3) the curve does not cross the zero phase shift axis at points greater than unity. Condition (2) is of particular interest because while it is judged stable on the basis of Nyquist's criterion it would appear to be unstable on the basis of the older transfer criterion discussed in the first and second sections.

For test purposes an amplifier was designed which, upon variation of an attenuator in the feed-back path, would satisfy each of the three above conditions in turn. The amplifier schematic is shown in Fig. 9.

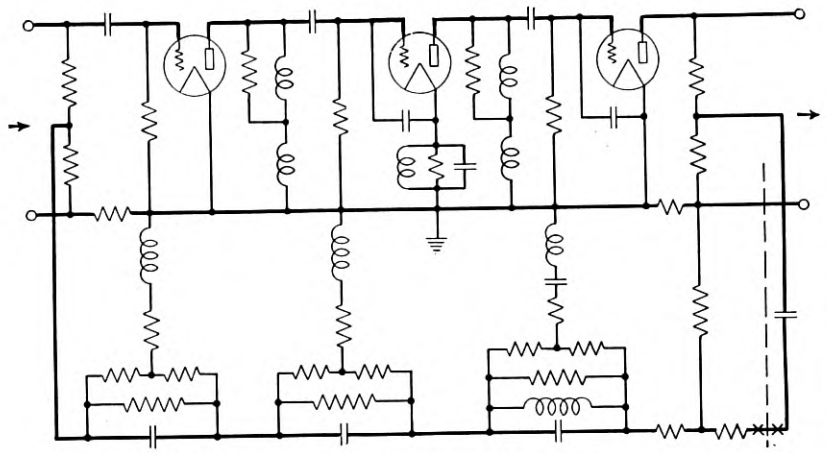


Fig. 9—Circuit diagram of the feed-back amplifier used in testing the stability criterion. The dashed line indicates the point at which the loop was broken for measurement of the transfer factor. At the left of this line is shown the resistance attenuator provided to vary the gain around the feed-back loop.

It has three stages, the first two tubes being space charge grid pentodes, and the last one a triode. The interstage coupling circuits were made up of simple inductances and resistances as shown. The amplifier was designed by E. L. Norton and E. E. Aldrich to provide a transfer factor characteristic having the desired shape, i.e., a loop crossing the zero phase axis in the neighborhood of 10 kilocycles. It will be observed that the feed-back circuit is connected between bridge networks in both input and output circuits, which were provided to eliminate reaction of the input and output circuits upon the feed-back network.[11]

Experimental Results

The transfer factor was measured for a zero setting of the feed-back attenuator over a frequency range of 0.5 to 1200 kilocycles. The results are shown in Fig. 10. The method of plotting this figure requires some discussion. In order to keep the curve within a reasonable size and still show the necessary details the scale has been made logarithmic by plotting the gain around the loop in decibels instead of the corresponding numerical ratios. It is of course impossible to carry this out completely on a polar diagram since the transfer factor goes to zero at high frequencies. To take care of this the scale is made logarithmic only above zero gain, corresponding to unit transfer ratio, and is linear below. It should be noted that if the logarithmic portion of the scale is translated outward so that the zero decibel point lies successively in the regions marked A, B, C, and D, the indicated amplifier conditions correspond to those designated above as (1), (2), (1) and (3) respectively. Experimentally an increase of the feed-back attenuator corresponds to such a translation of the logarithmic scale by an amount equal to the increase in attenuation. Therefore, the transition from one condition to another should occur when the attenuator setting is equal to the gain at a zero phase point in the curve as measured with a zero attenuator setting.

The test of the stability criterion consists of a determination of the attenuator settings at which oscillations begin, and a comparison of these settings with those at which a transition from a stable to an unstable condition is predicted by the theory. Experimentally oscillations were found to occur in regions A and C and not in regions B and D which is in qualitative agreement with Nyquist's predictions. Quantitatively the measured and predicted transition points agreed within one decibel which is estimated to be within the experimental error.

It should be noted that the plotted curve has been drawn up for $A(j\omega)$, no points of $A(-j\omega)$ being shown, although both are required

[11] H. S. Black, *Bell Sys. Tech. Jour.*, January, 1934.

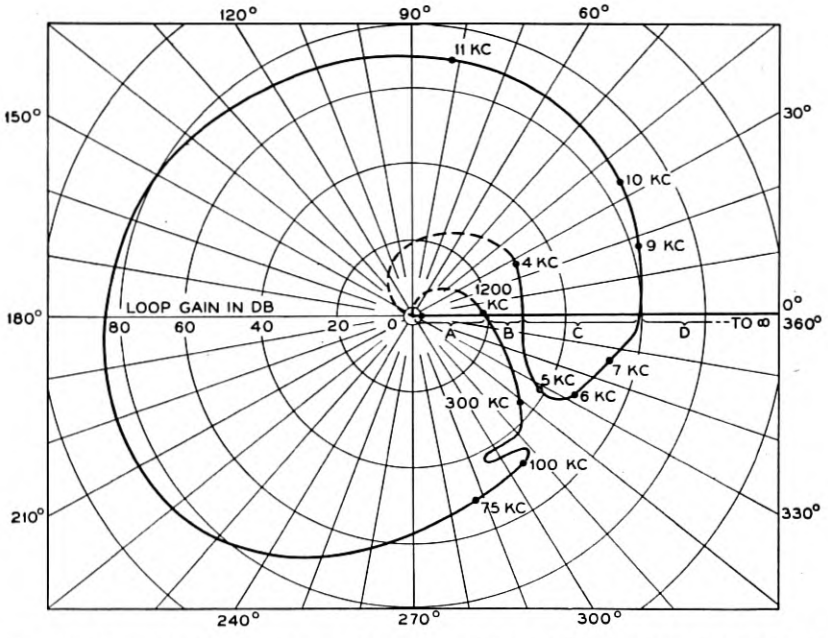


Fig. 10—Transfer factor diagram for the amplifier of Fig. 9 with the feed-back attenuator set at zero decibel.

by the theoretical derivation. Where the transfer factor is zero at zero frequency, only $A(j\omega)$ is required since the loop then closes for positive values of ω. In amplifiers transmitting d.c., however, both positive and negative values of ω are needed[8] to form a closed loop. In any case $A(-j\omega)$ is the mirror image of $A(j\omega)$ about the x-axis.

Extensions of the Criterion

Nonlinear Amplifier

The stability criterion which was verified by the experiments reported in the preceding section is framed for linear systems, those in which the steady state response is linearly proportional to the applied force. In vacuum tube circuits, linearity is best approximated at small force amplitudes, and is departed from to an extent dependent upon the impressed potentials, as well as upon tube and circuit characteristics. The divergence from linearity becomes well marked when the load capacities of the tubes are approached, or when grid current is made to flow through large grid impedances. The question then arises as to the form which the stability criterion takes when a tube circuit is

[8] Loc. cit.

operated in a nonlinear region—let us say by impressing upon the circuit a sufficiently large alternating potential provided by an external independent generator.

To answer this question we may consider the response of the amplifier, loaded by the independent generator, to a small alternating potential introduced for test purposes. Since the response of the system is known to be linear from the theory of perturbations, we might attempt to apply the linear criterion to the small superposed force. To do this it is necessary to measure the transfer factor for the small superposed force over the frequency range at a particular load of interest.

Application of the experimental technique to this extended criterion introduces difficulties since the opening of the feed-back loop for measuring purposes disturbs to a certain extent the distribution of these loads, particularly the harmonics, and modulation products in general. This makes it difficult to get the same loading effect when the loop is opened for measuring purposes as obtained when the loop is closed. Another consideration is that the response to the small component may be expected to vary in general at different points on the loading wave, so that the measuring procedure averages the response over a cycle of the loading wave. A method of measurement analogous to that of the flutter bridge would be required to evaluate the transfer factor at points of the loading cycle. Further, the measuring apparatus is affected by the presence of the loading currents when these are sufficiently large. In the present case in which the loading frequency (60 kilocycles) was far removed in the frequency scale from the test frequencies, it was found possible to approximate the necessary measurements by the insertion of selective circuits.

The curves of Fig. 11 represent portions of the transfer factor polar diagram for an amplifier similar to the one previously described, measured by the visual method with different loading amplitudes. The effect of the load on this particular amplifier is to change both phase shift and amplitude so that the curves shrink both radially and tangentially, pulling the loop back across the zero phase axis until, at the heaviest load, the two low-frequency crossings are completely eliminated. If the extended criterion is valid, we should expect the amplifier to be stable at any setting of the feed-back attenuator. As the load is decreased from this value, the crossings occur at successively higher gains so that the start of oscillations would occur at progressively higher settings of the feed-back attenuator.

The curves of Fig. 12 show the attenuator settings predicted by the extended criterion and those determined by direct observation of the attenuator setting required for oscillations when the feed-back circuit

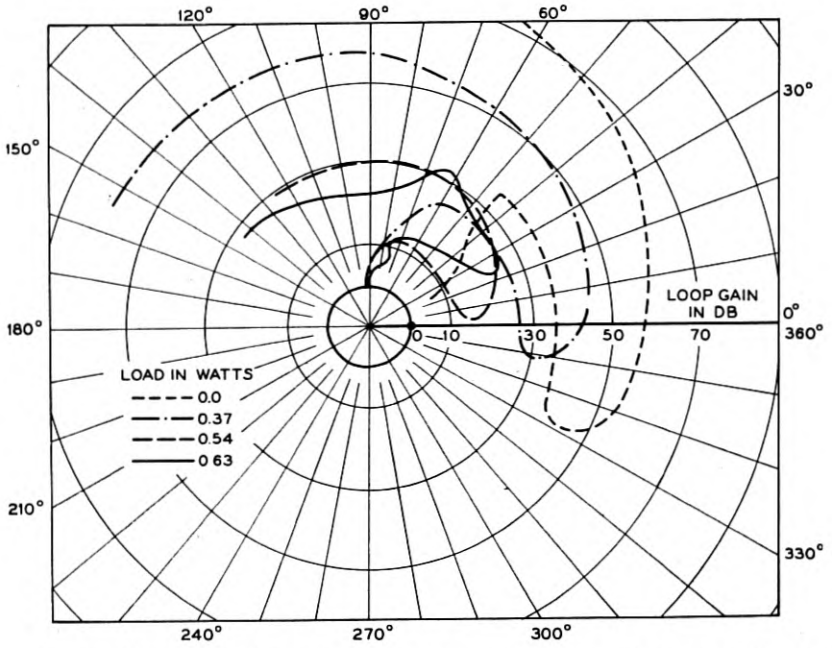


Fig. 11—The transfer factor diagram for the amplifier of Fig. 9 with the feed-back attenuator set at zero decibel. The four curves shown correspond to different amounts of the 60-kilocycle load.

was closed. Two sets of curves are shown, one for each of the low-frequency crossings. These are plotted against the loading amplitude. The agreement between the experimental and predicted values is close for the higher gain crossing at small loading amplitudes, but a divergence is apparent at high loads. For the lower crossing there is a divergence of 1.5 decibels at low loads, which changes sign and becomes greater at the higher loads. These divergences may be ascribed to a variety of causes among which probably the most important are the effects of harmonics upon the amplifier loading, overloading of the measuring apparatus by harmonics of the loading electromotive force, and phase shifts introduced by the selective circuits. The last two causes may be eliminated by improved technique, but the first cause in general introduces a fundamental difficulty, particularly important when large nonlinearities are involved.

Negative Impedances

One of the early forms of stability criterion mentioned in the first section was that relating to the measured impedance of the circuit.

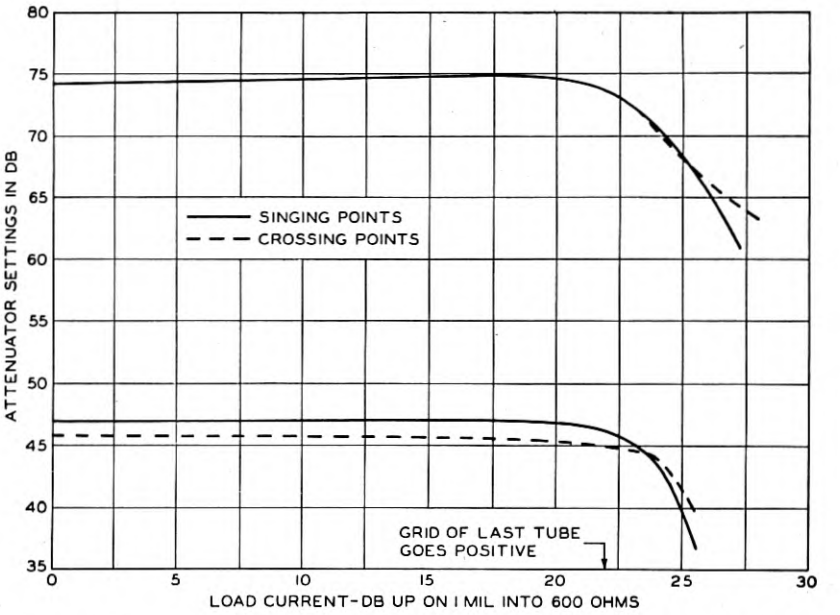


Fig. 12—Comparison of feed-back attenuator settings required for the starting of oscillations, and those deduced from the transfer diagram, plotted as functions of the 60-kilocycle load. The two dashed curves correspond to the two points (roughly 4.5 and 8 kilocycles) at which the transfer factor diagram (Fig. 11) crosses the zero phase axis. The gains of Figs. 11 and 12 cannot be compared directly because of a change made in the amplifier circuit of Fig. 12 which increased the loop gain.

Nyquist's criterion involving the transfer factor may be transformed so as to formulate a more complete criterion involving such an impedance.

To do this we have to express the factor $(1 - A)$, on which the stability criterion was based, in terms of the circuit impedances. For illustrative purposes we may quote the results obtained with the two fundamental forms of feed-back circuits, the series and shunt types.[12] These results, while obtained for the input circuit of the amplifier, are valid for any other point of the feed-back loop. Further, combinations of the shunt and series type feed-back circuits may be used.

Series Feed-Back

The series circuit is shown in Fig. 13, so called because the feed-back is applied in series with the amplied electromotive force and the amplifier input. The passive impedances marked are those existing when the feed-back loop is broken and terminated as indicated by the

[12] Crisson, *Bell Sys. Tech. Jour.*, vol. X, p. 485.

dotted lines. By direct circuit analysis, the current and voltage amplitudes in the feed-back condition are related by

$$E = (Z + Z_0 + Z_i)(1 - A)I,$$

where A and the Z's are functions of frequency. The total effective circuit impedance is obtained as the multiplifier of I in the right mem-

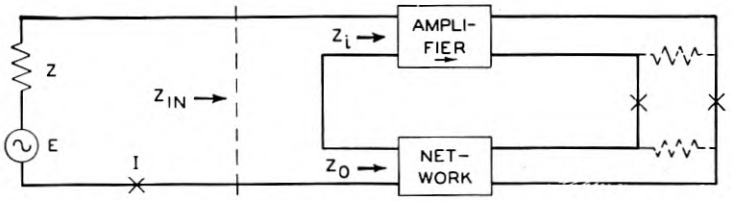


Fig. 13—Series type feed-back circuit. The dotted resistances indicate the terminations applied when the feed-back circuit was broken, to which the passive impedances (Z_iZ_0) apply. Z_{in} represents the effective input impedance with the feed-back circuit connected through.

ber. Subtracting the generator impedance Z from the total, the input impedance becomes

$$Z_{IN} = (Z_0 + Z_i)(1 - A) - AZ,$$

from which,

$$1 - A = \frac{Z}{Z + Z_0 + Z_i}\left(1 + \frac{Z_{IN}}{Z}\right).$$

Of the two factors of the right member, the first one, involving passive impedances alone, can have no roots with positive real part. Any such roots must, therefore, be contained in the second bracketed factor and then only when Z_{in} is negative. Hence paraphrasing the transfer factor criterion, if we plot $-Z_{in}/Z$ over the frequency range, the circuit is stable when the point (1, 0) is not enclosed by the resultant curve.

Shunt Feed-Back

Proceeding as in the series case with the circuit of Fig. 14 we get

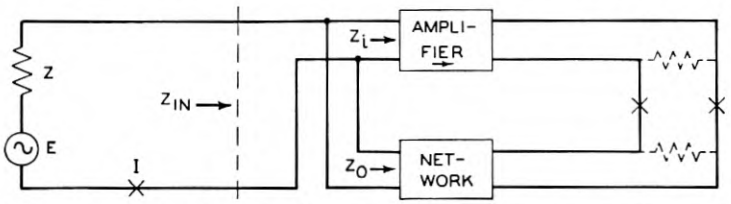


Fig. 14—Shunt type feed-back circuit. The notation corresponds to that of Fig. 13.

$$1 - A = \frac{Z_a}{Z + Z_a}\left(1 + \frac{Z}{Z_{in}}\right),$$

where Z_a represents the impedance of Z_0 and Z_i in parallel. Again only the bracketed term can yield undamped transients so that the criterion involves plotting $-Z/Z_{in}$ over the frequency range; if the resultant curve does not enclose (1, 0) the circuit is stable.

It may be remarked that these results are applicable to circuits including two-terminal negative impedances such as the oscillating arc and the dynatron, which are of the series and the shunt type respectively.

ACKNOWLEDGMENTS

The authors are indebted to Mr. L. W. Hussey for discussions of theoretical points, and to Mr. P. A. Reiling for his cooperation in the experiments.

Abstracts of Technical Articles from Bell System Sources

Shared Channel Broadcasting.[1] C. B. AIKEN. This paper deals with the experimental studies made on the character and causes of interference noticeable in shared channel broadcasting, such as heterodyning, flutter, sideband interference and wobbling. Valuable data are included on the characteristics of square-law and linear detectors anent to interference.

The Determination of Dielectric Properties at Very High Frequencies.[2] J. G. CHAFFEE. A simple method of determining the dielectric constant and power factor of solid dielectrics at frequencies as high as 20 megacycles, with an accuracy which is sufficient for most purposes, is described. The major sources of error are discussed in detail, and several precautions which should be observed are pointed out.

Measurements of the dielectric properties at 18 megacycles of a number of commonly used materials have shown that in general the power factor and dielectric constant are not widely different from those which obtain at frequencies of the order of one megacycle.

In addition, the results of an investigation of the input impedance of vacuum tube voltmeters at high frequencies are described as an illustration of the further application of this method of measurement.

Optical Factors in Caesium-Silver-Oxide Photoelectric Cells.[3] H. E. IVES AND A. R. OLPIN. This paper describes an investigation of the part played by the angle of incidence and state of polarization of the exciting light in producing the enhanced or selective emission of photoelectrons in the red region of the spectrum which is characteristic of photoelectric cells made by treating a silver surface with oxygen and caesium vapor (Fig. 1). This question is one which has been raised in connection with all types of photoelectric cells having composite surfaces and which exhibit spectrally selective emission. It has thus been an open question whether the selective peaks in the spectral response curves exhibited by the alkali hydride cells are to be ascribed to an enhanced effect of the perpendicular vector of obliquely incident

[1] *Radio Engineering*, June, 1934.
[2] *Proc. I. R. E.*, August, 1934.
[3] *Jour. Op. Soc. Am.*, August, 1934.

radiation, or whether the spectral selectivity is in the nature of a locally intrinsic emissive power, such as would be caused by an optical absorption band or an electronic transmission band. In order to answer this question, it is necessary to have emitting surfaces of a specular character. Such surfaces have not been prepared with the alkali hydrides, but it has been found possible to make the caesium-silver-oxide cells on specular plates of silver so that they retain their specular character in the final sensitized surface. Cells of this sort were used in this study, and have made possible a clear separation of the emissive singularities due to optical conditions and the singularities which may be described as intrinsic to the material.

Both from their method of preparation and from their optical behavior, we have felt justified in considering the caesium-silver-oxide photoelectric cells prepared with specular silver surfaces as consisting of silver surfaces overlaid with a thick layer of transparent refracting material, on the top of which is a thin photosensitive layer. The silver plates, after oxidation, exhibit interference colors, the exact color depending upon the amount of oxidation. Viewed at an angle through a nicol prism, these oxidized plates exhibit the well-known properties of thin refractive layers on a metal base. Thus when the plane of polarization is changed from the plane of incidence to the plane perpendicular thereto, no change of hue takes place for small angles of incidence; but at large angles, the color changes to a complementary hue. After the silver oxide surface has been exposed to caesium vapor and given a heat treatment, these optical properties are still usually observable, but degraded. The softening of the interference colors may be due either to a change in thickness of the refracting medium as caesium oxide is formed or to the introduction of a general body color. In a few less common cases the colors faded out completely, the plate at the end of the heat treatment being metallic in appearance yet still exhibiting a pronounced selective response to red and infrared light.

The behavior of a thin photoelectric sheet separated from a specular metal surface by a layer of refracting medium has been treated in an earlier paper where a layer of caesium was deposited on the top of a quartz-coated platinum plate. The data obtained in this earlier paper are immediately applicable to the present problem, granting the similarity of conditions which we have assumed. It has been convenient to pursue this present study on the assumption of such a similarity and to arrive at conclusions from the agreement with, or deviation from, the results obtained from the simpler materials and conditions previously studied.

Phase Angle of Vacuum Tube Transconductance at Very High Frequencies.[4] F. B. LLEWELLYN. Theoretical considerations indicate that the transconductance of a vacuum tube exhibits a phase angle when the transit time of electrons from cathode to anode becomes an appreciable fraction of the high-frequency period. Measurements show that such a phase angle actually occurs and that its behavior is in general agreement with the theoretical predictions.

Application of Sound Measuring Instruments to the Study of Phonetic Problems.[5] JOHN C. STEINBERG. This paper gives the results of a period by period analysis of the vowel sound waves occurring when the sentence "Joe took father's shoe bench out" was spoken. Such an analysis gives an approximate picture of the time variations in r.m.s. amplitude of the wave, frequency of voice fundamental, and frequency regions of overtone reenforcement. Although the study is confined to a few sounds and one speaker's voice, it illustrates a method of approach to studies of speech production and measurement.

[4] *Proc. I. R. E.*, August, 1934.
[5] *Jour. Acous. Soc. Am.*, July, 1934.

Contributors to this Issue

KARL K. DARROW, B.S., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., University of Chicago, 1917. Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Dr. Darrow has been engaged largely in writing on various fields of physics and the allied sciences.

LLOYD ESPENSCHIED. Mr. Espenschied is High Frequency Transmission Development Director in the Bell Telephone Laboratories. He joined the Bell System in 1910, having graduated from Pratt Institute the previous year. He has taken an important part in practically all of the Bell System radio developments, beginning with the first long-distance radio-telephone tests of 1915, at which time he received the voice in Hawaii from Arlington, Virginia. He has participated in a number of international conferences on electric communications.

J. G. KREER, B.S. in Electrical Engineering, University of Illinois, 1925; M.A., Columbia University, 1928. Bell Telephone Laboratories, 1925–. Mr. Kreer has been engaged in research work on carrier frequency systems.

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.

G. L. PEARSON, A.B., Willamette University, 1926; M.A. Stanford University, 1929. Bell Telephone Laboratories, 1929–. Mr. Pearson has been engaged in a study of the noise inherent in electric circuits.

D. B. PENICK, B.S. in Electrical Engineering, University of Texas, 1923; B.A., 1924; M.A. in Physics, Columbia University, 1927. Western Electric Company, Engineering Department, 1924–25; Bell Telephone Laboratories, 1925–. Mr. Penick has been engaged in special problems related to the development of vacuum tubes.

E. PETERSON, Cornell University, 1911–14; Brooklyn Polytechnic, E.E., 1917; Columbia, A.M., 1923; Ph.D., 1926; Electrical Testing Laboratories, 1915–17; Signal Corps, U. S. Army, 1917–19. Bell Telephone Laboratories, 1919–. Dr. Peterson's work has been largely in theoretical studies of carrier current apparatus.

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.

S. A. SCHELKUNOFF, B.A., M.A., in Mathematics, The State College of Washington, 1923; Ph.D. in Mathematics, Columbia University, 1928. Engineering Department, Western Electric Company, 1923–25. Bell Telephone Laboratories, 1925–26. Department of Mathematics, State College of Washington, 1926–29. Bell Telephone Laboratories, 1929–. Dr. Schelkunoff has been engaged in mathematical research, especially in the field of electromagnetic theory.

M. E. STRIEBY, A.B., Colorado College, 1914; B.S., Harvard, 1916; B.S. in E.E., M.I.T., 1916; New York Telephone Company, Engineering Department, 1916–17; Captain, Signal Corps, U. S. Army, A. E. F., 1917–19. American Telephone and Telegraph Company, Department of Development and Research, 1919–29; Bell Telephone Laboratories, 1929–. Mr. Strieby has been associated with various phases of transmission work, more particularly with the development of long toll circuits. At the present time, in his capacity as Carrier Transmission Research Engineer, he directs studies of new and improved methods of carrier frequency transmission over existing or new facilities.

L. A. WARE, B.E., Engineering College, University of Iowa, 1926; M.S., University of Iowa, 1927; PhD., Physics Department, University of Iowa, 1930. Instructor in Physics, University of Iowa, 1926–29. Bell Telephone Laboratories, 1929–. Dr. Ware's work has been chiefly in connection with regenerative amplifier development.