The Bell System Technical Journal

Vol. XIX *July, 1940* *No. 3*

Crosstalk in Coaxial Cables—Analysis Based on Short-Circuited and Open Tertiaries

By K. E. GOULD

The problem considered herein is that of estimating, from measurements on short lengths of coaxial cable, the crosstalk to be expected in long lengths of the same cable. The method developed, which is particularly applicable to cases in which the effect of tertiary circuits on the crosstalk is large, is based on measurements of crosstalk in a short length, with the tertiaries first short-circuited and then open. The application of this method to the cable described in the companion paper by Messrs. Booth and Odarenko gave crosstalk values in good agreement with their experimental results.

Introduction

FOR a number of years the problem of crosstalk summation in long open-wire lines or cables has been studied by measuring crosstalk, in phase and magnitude, in short lengths. The crosstalk within a short length, between two circuits terminated in their characteristic impedances, would be measured with all important tertiary circuits also approximately terminated. Then the crosstalk between two circuits in adjoining short lengths would be measured with the tertiary circuits terminated. From these coefficient measurements the crosstalk in a long length could be estimated by a process of integration.

The application of this general method to crosstalk in the usual types of coaxial cable would require great accuracy in the coefficient measurements, because in longer lengths the desired crosstalk value depends on the difference between two nearly equal quantities involving the coefficients. In the following analysis the computation of the crosstalk for long lengths of coaxial cable is based on crosstalk measurements, in phase and magnitude, between two coaxial circuits in a single short length with the tertiary circuits first open and then short-circuited, no crosstalk measurements with terminated tertiary circuits being involved.

This method of analysis, when applied to the twin coaxial cable described in the companion paper by Messrs. Booth and Odarenko, gave results in good agreement with the measured crosstalk.

In this analysis it was assumed that all the tertiary circuits could be combined and considered as a single circuit. Although no evidence has been found that with the types of structure studied so far, better accuracy would result from the further refinement of considering two or more dissimilar tertiary circuits with coupling between them, there is one case of practical importance which cannot be handled with the single-tertiary analysis. This case is that of the interaction crosstalk (that is, the crosstalk by way of a tertiary circuit) between two adjoining lengths of coaxial cable, when, at the junction, part of the tertiary conductors are short-circuited to the outer coaxial conductors while the remaining tertiary conductors continue through with no discontinuity. This problem might be of importance where, at a repeater, the outer coaxial conductors and the sheath are bonded together, but paper-insulated pairs in the same sheath provide an uninterrupted tertiary circuit. The near-end crosstalk under such conditions might also differ significantly from the values indicated by the single-tertiary analysis.

The two-tertiary analysis is too long to be given here in detail, and hence has been outlined only to such an extent as to indicate the derivation of the formulas for interaction crosstalk when one of the tertiaries is short circuited and the other terminated in its characteristic impedance. The formula for near-end crosstalk under this condition is given without derivation.

I—Identical Coaxial Lines Symmetrically Placed with Respect to a Single Tertiary

The first case we shall consider herein is that of any number of identical coaxial lines with the outer coaxial conductors in continuous electrical contact and symmetrically placed with respect to a single tertiary circuit, such as that which might be provided by a sheath surrounding the coaxial lines and insulated from the outer coaxial conductors, or by a surrounding layer of paper-insulated pairs. Throughout we shall assume that the reaction of the induced currents upon the disturbing line is negligible.

Following a nomenclature analogous to that of the Schelkunoff-Odarenko paper,[1] we will designate by Z_{12} the mutual impedance per unit length between any two coaxial lines in the presence of the other coaxial lines but in the absence of any other conductors. The mutual

[1] *Bell System Technical Journal*, April, 1937.

impedance per unit length between any coaxial line (in the presence of the other coaxial lines) and the tertiary circuit consisting of all of the coaxial outer conductors with return by way of the sheath or other tertiary conductors, we will designate by Z_{13}.

If we consider the crosstalk between two coaxial lines of length, l, such that the coaxial lines and the tertiary are electrically short, each coaxial line being terminated in its characteristic impedance Z and the tertiary open at each end, the crosstalk (near-end and far-end being identical for such a length) is given by

$$\frac{Z_{12}l}{2Z}.$$

If, now, we consider a case similar except that the tertiary is short-circuited at each end, the crosstalk is the above term plus the effect of the tertiary current I_3, which, for unit current in the disturbing coaxial line, is given by

$$I_3 = \frac{Z_{13}}{Z_{33}},$$

where Z_{33} is the series impedance of the tertiary circuit per unit length. This tertiary current will produce a current $\left(-\frac{Z^2_{13}l}{2ZZ_{33}} \right)$ in the disturbed coaxial line and the total crosstalk will be

$$\frac{Z_{12}l}{2Z} - \frac{Z^2_{13}l}{2ZZ_{33}}.$$

If we designate $\frac{Z_{12}}{2Z}$ by X and $\frac{Z^2_{13}}{Z_{12}Z_{33}}$ by ξ, then, for an electrically short length, X will represent the crosstalk per unit length between two coaxial lines with the tertiary open, and $X(1 - \xi)$ the crosstalk per unit length with the tertiary short-circuited. In the formulas developed below these quantities will be found to be of fundamental importance.

Tertiary Terminated in its Characteristic Impedance

Far-End Crosstalk

From the Schelkunoff-Odarenko paper, the sum of the direct far-end crosstalk (eq. 19) and the indirect far-end crosstalk (eq. 40) for any length under these conditions gives the total far-end crosstalk F_t, as

$$F_t = \frac{Z_{12}l}{2Z} - \frac{Z^2_{13}}{4ZZ_3}\left[\frac{2\gamma_3 l}{\gamma_3^2 - \gamma^2} - \frac{1 - e^{-(\gamma_3 - \gamma)l}}{(\gamma_3 - \gamma)^2} - \frac{1 - e^{-(\gamma_3 + \gamma)l}}{(\gamma_3 + \gamma)^2} \right], \quad (1)$$

where

Z_3 = characteristic impedance of tertiary circuit,

γ, γ_3 = propagation constants of coaxial lines and tertiary circuit respectively.

This may be rearranged and written (since $Z_3\gamma_3 = Z_{33}$)

$$F_t = \frac{Z_{12}l}{2Z} - \frac{Z^2_{13}l}{2ZZ_{33}} - \frac{Z^2_{13}}{2ZZ_{33}} \times \left[\frac{l\gamma^2}{\gamma_3^2 - \gamma^2} - \frac{\gamma_3}{2}\left(\frac{1 - e^{-(\gamma_3-\gamma)l}}{(\gamma_3 - \gamma)^2} + \frac{1 - e^{-(\gamma_3+\gamma)l}}{(\gamma_3 + \gamma)^2} \right) \right] \quad (2a)$$

$$= X\left[l(1 - \xi) - l\xi\frac{\gamma^2}{\gamma_3^2 - \gamma^2} + \frac{\xi\gamma_3}{2} \times \left(\frac{1 - e^{-(\gamma_3-\gamma)l}}{(\gamma_3 - \gamma)^2} + \frac{1 - e^{-(\gamma_3+\gamma)l}}{(\gamma_3 + \gamma)^2} \right) \right]. \quad (2b)$$

This formula has been found to be applicable, with good accuracy, to the types of coaxial cable which have been studied so far. The quantities X and $X(1 - \xi)$ are determined from crosstalk measurements on a short length, and the propagation constants are of course readily determined.

Near-End Crosstalk

A similar approach to the problem of the near-end crosstalk N_t with the tertiary terminated in its characteristic impedance, using equations (10) and (32) of the Schelkunoff-Odarenko paper, gives

$$N_t = X\left[(1 - e^{-2\gamma l}) \left(\frac{1 - \xi}{2\gamma} - \frac{\xi\gamma}{2(\gamma_3^2 - \gamma^2)} \right) + \frac{\xi\gamma_3}{2(\gamma_3^2 - \gamma^2)} (1 + e^{-2\gamma l} - 2e^{-(\gamma_3+\gamma)l}) \right]. \quad (3)$$

Here, as in the case of equation (2b) above, the crosstalk may be computed readily from crosstalk and impedance measurements on a short sample.

Interaction Crosstalk

Far-End Far-End and Far-End Near-End

We will consider the interaction crosstalk between two adjoining sections of lengths l and l', respectively, the tertiary being connected through at the junction, with no discontinuity. The tertiary current $i_3(l)$ at the far end of a section of length l, for unit sending-end current, with the tertiary terminated in its characteristic impedance, is readily formulated as

$$i_3(l) = \frac{Z_{13}}{2Z_3} \frac{e^{-\gamma l} - e^{-\gamma_3 l}}{\gamma_3 - \gamma}. \quad (4)$$

In the adjoining section, with the tertiary terminated in its characteristic impedance, the tertiary current $i_3(y)$ will be given by $i_3(l)e^{-\gamma_3 y}$, where y is the distance measured from the junction of the two sections.

This tertiary current $i_3(y)$ will produce a far-end current in the disturbed coaxial of

$$\frac{Z^2_{13}}{4ZZ_3}\frac{e^{-\gamma l} - e^{-\gamma_3 l}}{\gamma_3 - \gamma}\int_0^{l'} e^{-\gamma_3 y}e^{-\gamma(l'-y)}dy.$$

The equal-level far end-far end interaction crosstalk FF, being this far-end current divided by $e^{-\gamma(l+l')}$, may be obtained as

$$FF = \frac{X\xi\gamma_3}{2(\gamma_3 - \gamma)^2}(1 - e^{-(\gamma_3-\gamma)l})(1 - e^{-(\gamma_3-\gamma)l'}). \tag{5}$$

The near-end current in the disturbed coaxial due to the current $i_3(y)$ is given by

$$\frac{Z^2_{13}}{4ZZ_3}\frac{e^{-\gamma l} - e^{-\gamma_3 l}}{\gamma_3 - \gamma}\int_0^{l'} e^{-\gamma_3 y}e^{-\gamma y}dy.$$

From this the equal-level far end-near end interaction crosstalk FN, being this near-end current divided by $e^{-\gamma l}$, may be obtained as

$$FN = \frac{X\xi\gamma_3}{2(\gamma_3^2 - \gamma^2)}(1 - e^{-(\gamma_3-\gamma)l})(1 - e^{-(\gamma_3+\gamma)l'}). \tag{6}$$

Near-End Near-End

The near-end tertiary current in the section of length l is similarly formulated as

$$i_3(0) = \frac{Z_{13}}{2Z_3}\frac{1 - e^{-(\gamma_3+\gamma)l}}{\gamma_3 + \gamma}. \tag{7}$$

The near-end near-end interaction crosstalk NN is readily obtained, in a fashion similar to that outlined above for the far-end near-end interaction crosstalk, as

$$NN = \frac{X\xi\gamma_3}{2(\gamma_3 + \gamma)^2}(1 - e^{-(\gamma_3+\gamma)l})(1 - e^{-(\gamma_3+\gamma)l'}). \tag{8}$$

Tertiary Short-Circuited

The general case of the crosstalk between coaxial lines of length l with the tertiary short-circuited at each end may be attacked as follows. At any point at a distance x from the sending end, the voltage gradient along the outer surface of the outer coaxial conductors, for unit sending-end current, will be $Z_{13}e^{-\gamma x}$. Each differential element, $Z_{13}e^{-\gamma x}dx$, of this voltage drop will produce a current in the tertiary circuit de-

termined by the impedances, Z' and Z'', of the tertiary as seen in the two directions from this point, these impedances being

$$Z' = Z_3 \tanh \gamma_3 x \tag{9}$$

and

$$Z'' = Z_3 \tanh \gamma_3 (l - x) \tag{10}$$

respectively toward and away from the sending end.

At any other point at a distance y from the sending end, the tertiary current due to the voltage $Z_{13}e^{-\gamma x}dx$ will be given, for $y > x$, by

$$i_3(y) = \frac{Z_{13}e^{-\gamma x}dx}{Z' + Z''} \frac{\cosh \gamma_3 (l - y)}{\cosh \gamma_3 (l - x)}. \tag{11}$$

From this the transfer admittance $A(x, y)$ between these two points is obtained as

$$A(x, y) = \frac{1}{Z_3} \frac{\cosh \gamma_3 (l - y)}{\tanh \gamma_3 x \cosh \gamma_3 (l - x) + \sinh \gamma_3 (l - x)}. \tag{12}$$

Similarly, for $y < x$, this transfer admittance is obtained as

$$A(x, y) = \frac{1}{Z_3} \frac{\cosh \gamma_3 y}{\sinh \gamma_3 x + \cosh \gamma_3 x \tanh \gamma_3 (l - x)}. \tag{13}$$

The tertiary current, $i_3(x)$, is given by

$$i_3(x) = \int_0^l Z_{13}e^{-\gamma y}A(x, y)dy \tag{14}$$

from which we obtain

$$i_3(x) = \frac{Z_{13}}{2Z_3 \sinh \gamma_3 l} \times \left[\begin{array}{l} \dfrac{1}{\gamma_3 + \gamma} \left[\begin{array}{r} \cosh \gamma_3 (l - x) + e^{-\gamma x} \sinh \gamma_3 l \\ - e^{-\gamma l} \cosh \gamma_3 x \end{array} \right] \\ - \dfrac{1}{\gamma_3 - \gamma} \left[\begin{array}{r} \cosh \gamma_3 (l - x) - e^{-\gamma x} \sinh \gamma_3 l \\ - e^{-\gamma l} \cosh \gamma_3 x \end{array} \right] \end{array} \right]. \tag{15}$$

Far-End Crosstalk

The indirect far-end crosstalk F_s' due to this tertiary current (eq. 15) is given by

$$F_s' = e^{\gamma l} \int_0^l \frac{Z_{13} i_3(x)}{2Z} e^{-\gamma (l - x)} dx \tag{16a}$$

$$= \frac{Z^2_{13}}{2ZZ_{33}} \left[\frac{l\gamma_3^2}{\gamma_3^2 - \gamma^2} - \frac{2\gamma_3\gamma^2}{(\gamma_3^2 - \gamma^2)^2} \frac{\cosh \gamma_3 l - \cosh \gamma l}{\sinh \gamma_3 l} \right]. \tag{16b}$$

If this is combined with the direct far-end crosstalk $\frac{Z_{12}l}{2Z}$, and the terms rearranged as in the case of equation (2b), the total far-end crosstalk F_s is obtained as

$$F_s = X\left[l(1-\xi) - l\xi\frac{\gamma^2}{\gamma_3^2-\gamma^2} + 2\xi\frac{\gamma_3\gamma^2}{(\gamma_3^2-\gamma^2)^2}\frac{\cosh\gamma_3 l - \cosh\gamma l}{\sinh\gamma_3 l}\right]. \quad (17)$$

It will be noted that equations (2b) and (17) differ only in the terms which are not proportional to the length and which thus are of decreasing importance as the length becomes great.

Near-End Crosstalk

The indirect near-end crosstalk N_s' due to the tertiary current $i_3(x)$ is given by

$$N_s' = \int_0^l \frac{Z_{13}i_3(x)}{2Z}e^{-\gamma x}dx \quad (18)$$

By substituting $i_3(x)$ from equation (15) herein, and combining the result with the direct near-end crosstalk,

$$\frac{Z_{12}}{2Z}\frac{1-e^{-2\gamma l}}{2\gamma}$$

we obtain the total near-end crosstalk N_s, which may be written in the form

$$N_s = X\left[\frac{1-e^{-2\gamma l}}{2\gamma}\left((1-\xi)+\frac{\xi\gamma^2(\gamma_3^2+\gamma^2)}{(\gamma_3^2-\gamma^2)^2}\right) - \left(\frac{\xi\gamma_3\gamma^2}{(\gamma_3^2-\gamma^2)^2}\right)\left(\frac{(1+e^{-2\gamma l})\cosh\gamma_3 l - 2e^{-\gamma l}}{\sinh\gamma_3 l}\right)\right]. \quad (19)$$

II—Identical Coaxial Lines Symmetrically Placed with Respect to Each of Two Dissimilar Tertiaries

We will now consider the case of any number of identical coaxial lines with the outer conductors in continuous electrical contact and symmetrically placed with respect to each of two dissimilar tertiaries with coupling between them.

In an unpublished memorandum by J. Riordan, the general forms are developed for the currents and voltages in two parallel circuits having uniformly distributed self and mutual impedances and admittances, when these circuits are subjected to impressed axial fields.

These currents (I_1 and I_2) and voltages (V_1 and V_2) (the subscripts applying, of course, to the respective tertiaries) are given by the coefficient array,

	$(a_1 + P_1)e^{-\gamma_1 x}$	$(b_1 + Q_1)e^{\gamma_1 x}$	$(a_2 + P_2)e^{-\gamma_2 x}$	$(b_2 + Q_2)e^{\gamma_2 x}$
I_1	1	-1	η_2	$-\eta_2$
I_2	η_1	$-\eta_1$	1	-1
V_1	K_1	K_1	$-\eta_1 K_2$	$-\eta_1 K_2$
V_2	$-\eta_2 K_1$	$-\eta_2 K_1$	K_2	K_2

where a_1, b_1, a_2 and b_2 are constants to be determined from the boundary conditions, and

$$\begin{matrix} P_1 \\ Q_1 \end{matrix} = \frac{1}{2K_1(1 - \eta_1\eta_2)} \int e^{\pm\gamma_1 x}(f_1 + \eta_1 f_2)dx, \tag{20}$$

$$\begin{matrix} P_2 \\ Q_2 \end{matrix} = \frac{1}{2K_2(1 - \eta_1\eta_2)} \int e^{\pm\gamma_2 x}(\eta_2 f_1 + f_2)dx, \tag{21}$$

f_1 and f_2 being the impressed fields along circuits 1 and 2 respectively.

If we consider the two tertiary circuits as consisting of (1) the outer coaxial conductors in parallel with return by tertiary path 1 and (2) tertiary path 1 − tertiary path 2, only tertiary circuit 1 will be subjected to an impressed field. Thus we will have $f_2 = 0$ and $f_1 = Z_{13}e^{-\gamma x}$ (for unit sending-end current in the disturbing coaxial line), where Z_{13} is the mutual impedance, per unit length, between a coaxial (in the presence of the other paralleling coaxials) and tertiary 1, and γ is the propagation constant, per unit length, for the coaxial circuit. The other quantities in this array are circuit parameters given as follows in terms of the series impedances Z_{11}, Z_{22} and Z_{12} per unit length and admittances a_{11}, a_{22} and a_{12} per unit length (subscripts 11 and 22 for self impedance or self admittance of circuits 1 and 2 respectively, and 12 for mutuals):

$$\begin{matrix} \gamma_1^2 \\ \gamma_2^2 \end{matrix} = \tfrac{1}{2}[a_{11}Z_{11} + a_{22}Z_{22} + 2a_{12}Z_{12} \pm ((a_{11}Z_{11} - a_{22}Z_{22})^2 + 4(a_{11}Z_{12} + a_{12}Z_{22})(a_{12}Z_{11} + a_{22}Z_{12}))^{1/2}], \tag{22}$$

$$\eta_1 = \frac{\gamma_1^2 - a_{11}Z_{11} - a_{12}Z_{12}}{a_{11}Z_{12} + a_{12}Z_{22}}, \tag{23}$$

$$\eta_2 = \frac{\gamma_2^2 - a_{12}Z_{12} - a_{22}Z_{22}}{a_{12}Z_{11} + a_{22}Z_{12}}, \tag{24}$$

$$K_1 = \frac{Z_{11} + \eta_1 Z_{12}}{\gamma_1} = \frac{\gamma_1}{a_{11} - \eta_2 a_{12}}, \tag{25}$$

$$K_2 = \frac{Z_{22} + \eta_2 Z_{12}}{\gamma_2} = \frac{\gamma_2}{a_{22} - \eta_1 a_{12}}. \tag{26}$$

From equations (20) and (21) above, we have

$$P_1 = \frac{Z_{13}}{2K_1(1 - \eta_1\eta_2)(\gamma_1 - \gamma)} e^{(\gamma_1 - \gamma)x}, \tag{27}$$

$$Q_1 = \frac{-Z_{13}}{2K_1(1 - \eta_1\eta_2)(\gamma_1 + \gamma)} e^{-(\gamma_1 + \gamma)x}, \tag{28}$$

$$P_2 = \frac{Z_{13}\eta_2}{2K_2(1 - \eta_1\eta_2)(\gamma_2 - \gamma)} e^{(\gamma_2 - \gamma)x}, \tag{29}$$

$$Q_2 = \frac{-Z_{13}\eta_2}{2K_2(1 - \eta_1\eta_2)(\gamma_2 + \gamma)} e^{-(\gamma_2 + \gamma)x}. \tag{30}$$

If we designate

$$\frac{Z_{13}}{K_1(1 - \eta_1\eta_2)(\gamma_1^2 - \gamma^2)} \text{ by } \psi_1$$

and

$$\frac{Z_{13}\eta_2}{K_2(1 - \eta_1\eta_2)(\gamma_2^2 - \gamma^2)} \text{ by } \psi_2,$$

we have

$$I_1 = a_1e^{-\gamma_1 x} - b_1e^{\gamma_1 x} + \eta_2 a_2 e^{-\gamma_2 x} - \eta_2 b_2 e^{\gamma_2 x} + (\psi_1\gamma_1 + \psi_2\eta_2\gamma_2)e^{-\gamma x}, \tag{31}$$

$$I_2 = \eta_1 a_1 e^{-\gamma_1 x} - \eta_1 b_1 e^{\gamma_1 x} + a_2 e^{-\gamma_2 x} - b_2 e^{\gamma_2 x} + (\psi_1\eta_1\gamma_1 + \psi_2\gamma_2)e^{-\gamma x}, \tag{32}$$

$$V_1 = K_1a_1e^{-\gamma_1 x} + K_1b_1e^{\gamma_1 x} - \eta_1 K_2 a_2 e^{-\gamma_2 x} - \eta_1 K_2 b_2 e^{\gamma_2 x} + (\psi_1 K_1\gamma - \psi_2 K_2\eta_1\gamma)e^{-\gamma x}, \tag{33}$$

$$V_2 = -\eta_2 K_1 a_1 e^{-\gamma_1 x} - \eta_2 K_1 b_1 e^{\gamma_1 x} + K_2 a_2 e^{-\gamma_2 x} + K_2 b_2 e^{\gamma_2 x} - (\psi_1 K_1\eta_2\gamma - \psi_2 K_2\gamma)e^{-\gamma x}. \tag{34}$$

Before proceeding with the application of these results to specific crosstalk problems, we will establish certain relations which, as in the single-tertiary analysis, will be fundamental in relating crosstalk measurements on short lengths of cable to the crosstalk to be expected in a longer length.

Let us consider the crosstalk as measured on a short length under the following two conditions: (1) both tertiaries open and (2) tertiary 1 short-circuited at each end and tertiary 2 open. We will designate the crosstalk under condition (1) by Xl and under condition (2) by $Xl(1 - \xi)$. Under condition (2) the tertiary current (I_1) for unit current in the energized coaxial is given by $\frac{Z_{13}}{Z_{11}}$ and the indirect crosstalk current in the disturbed coaxial is thus $\frac{-Z^2_{13}l}{2ZZ_{11}}$, so that we have

$$X\xi = \frac{Z^2_{13}}{2ZZ_{11}}. \tag{35}$$

Interaction Crosstalk with One Tertiary Short-Circuited Far-End

For the sake of simplicity, and with no considerable loss of applicability, we will postulate the restriction that $e^{\gamma_1 l}$ and $e^{\gamma_2 l}$ are large compared with $e^{\gamma l}$, where l is the length of the section in which we are formulating the tertiary currents.

Referring to equations (31) to (34), under the above restrictions the terms involving $e^{-\gamma_1 x}$ and $e^{-\gamma_2 x}$ are negligible in the region near $x = l$ and thus in this region

$$I_1 = -b_1 e^{\gamma_1 x} - \eta_2 b_2 e^{\gamma_2 x} + [(p_1 - q_1) + \eta_2(p_2 - q_2)]e^{-\gamma x}, \tag{36}$$

$$I_2 = -\eta_1 b_1 e^{\gamma_1 x} - b_2 e^{\gamma_2 x} + [\eta_1(p_1 - q_1) + (p_2 - q_2)]e^{-\gamma x}, \tag{37}$$

$$V_1 = K_1 b_1 e^{\gamma_1 x} - \eta_1 K_2 b_2 e^{\gamma_2 x} + [K_1(p_1 + q_1) - \eta_1 K_2(p_2 + q_2)]e^{-\gamma x}, \tag{38}$$

$$V_2 = -\eta_2 K_1 b_1 e^{\gamma_1 x} + K_2 b_2 e^{\gamma_2 x} + [-\eta_2 K_1(p_1 + q_1) + K_2(p_2 + q_2)]e^{-\gamma x}, \tag{39}$$

where [2]

$$p_1 = \frac{Z_{13}}{2K_1(1 - \eta_1\eta_2)(\gamma_1 - \gamma)}, \tag{40}$$

$$q_1 = \frac{-Z_{13}}{2K_1(1 - \eta_1\eta_2)(\gamma_1 + \gamma)}, \tag{41}$$

$$p_2 = \frac{Z_{13}\eta_2}{2K_2(1 - \eta_1\eta_2)(\gamma_2 - \gamma)}, \tag{42}$$

$$q_2 = \frac{-Z_{13}\eta_2}{2K_2(1 - \eta_1\eta_2)(\gamma_2 + \gamma)}. \tag{43}$$

If, now, x is measured from the far end, and the following substitutions are made as a matter of convenience: [3]

$$a_1 = b_1 e^{(\gamma_1 + \gamma)l}, \tag{44}$$

$$a_2 = b_2 e^{(\gamma_2 + \gamma)l}, \tag{45}$$

equations (36) to (39), multiplied by $e^{\gamma l}$ so that the currents and voltages are given for unit received current in the energized coaxial, become

$$I_1 = -a_1 e^{-\gamma_1 x} - \eta_2 a_2 e^{-\gamma_2 x} + [(p_1 - q_1) + \eta_2(p_2 - q_2)]e^{\gamma x}, \tag{46}$$

$$I_2 = -\eta_1 a_1 e^{-\gamma_1 x} - a_2 e^{-\gamma_2 x} + [\eta_1(p_1 - q_1) + (p_2 - q_2)]e^{\gamma x}, \tag{47}$$

[2] The terms involving $e^{-\gamma x}$ here are identical with the corresponding terms in equations (31) to (34) except for the change in nomenclature, which in each case has been chosen so that the a's and b's will be given by simple functions of the parameters employed (p's and q's here: ψ's in the previous equations).

[3] a_1 and a_2 here have no relation to a_1 and a_2 in equations (31) to (34).

$$V_1 = K_1a_1e^{-\gamma_1 x} - \eta_1K_2a_2e^{-\gamma_2 x} + [K_1(p_1 + q_1) - \eta_1K_2(p_2 + q_2)]e^{\gamma x}, \quad (48)$$

$$V_2 = -\eta_2K_1a_1e^{-\gamma_1 x} + K_2a_2e^{-\gamma_2 x} + [-\eta_2K_1(p_1 + q_1) + K_2(p_2 + q_2)]e^{\gamma x}. \quad (49)$$

In the section, of length l', adjacent to the far end of the energized section, the impressed fields are zero and thus (under the condition that $e^{\gamma_1 l'}$ and $e^{\gamma_2 l'}$ are large compared with unity), using primes to indicate currents and voltages in this region, with the distance x' taken positive from $x = l$,

$$I_1' = a_1'e^{-\gamma_1 x'} + \eta_2a_2'e^{-\gamma_2 x'}, \quad (50)$$

$$I_2' = \eta_1a_1'e^{-\gamma_1 x'} + a_2'e^{-\gamma_2 x'}, \quad (51)$$

$$V_1' = K_1a_1'e^{-\gamma_1 x'} - \eta_1K_2a_2'e^{-\gamma_2 x'}, \quad (52)$$

$$V_2' = -\eta_2K_1a_1'e^{-\gamma_1 x'} + K_2a_2'e^{-\gamma_2 x'}. \quad (53)$$

With tertiary 1 short-circuited, the boundary conditions to be satisfied are that at $x = x' = 0$, $V_1 = V_1' = 0$ and $I_2 = I_2'$. From these boundary conditions, we obtain

$$a_1 = -(p_1 + q_1) + \frac{\eta_1K_2(\eta_1p_1 + p_2)}{K_1 + \eta_1^2K_2}, \quad (54)$$

$$a_2 = -(p_2 + q_2) + \frac{K_1(\eta_1p_1 + p_2)}{K_1 + \eta_1^2K_2}, \quad (55)$$

$$a_1' = \frac{\eta_1K_2(\eta_1p_1 + p_2)}{K_1 + \eta_1^2K_2}, \quad (56)$$

$$a_2' = \frac{K_1(\eta_1p_1 + p_2)}{K_1 + \eta_1^2K_2}. \quad (57)$$

The equal-level far-end far-end interaction crosstalk FF_s is given by [4]

$$FF_s = \frac{Z_{13}}{2Z}e^{\gamma l'}\int_0^{l'} I_1'e^{-\gamma(l'-x')}dx'. \quad (58)$$

With I_1' as given by equations (50), (56) and (57), under the restrictions we have placed on $\gamma_1 l'$ and $\gamma_2 l'$, we have

$$FF_s = \frac{Z^2_{13}}{4Z(1 - \eta_1\eta_2)(K_1 + \eta_1^2K_2)} \times \left[\frac{\eta_1}{K_1(\gamma_1 - \gamma)} + \frac{\eta_2}{K_2(\gamma_2 - \gamma)}\right]\left[\frac{\eta_1K_2}{\gamma_1 - \gamma} + \frac{\eta_2K_1}{\gamma_2 - \gamma}\right] \quad (59)$$

[4] As pointed out in the Schelkunoff-Odarenko paper in the section on mutual impedance, since a coaxial circuit is involved, the current distribution external to this circuit does not affect the mutual impedance, and hence the current I_2' contributes nothing to the crosstalk.

or, with the use of equation (35),

$$FF_s = \frac{X\xi Z_{11}}{2(1-\eta_1\eta_2)(K_1+\eta_1^2K_2)} \times \left[\frac{\eta_1}{K_1(\gamma_1-\gamma)}+\frac{\eta_2}{K_2(\gamma_2-\gamma)}\right]\left[\frac{\eta_1K_2}{\gamma_1-\gamma}+\frac{\eta_2K_1}{\gamma_2-\gamma}\right]. \quad (60)$$

The equal-level far-end near-end interaction crosstalk FN_s is given by

$$FN_s = \frac{Z_{13}}{2Z}\int_0^{l'} I_1' e^{-\gamma x'} dx', \quad (61)$$

and under the restrictions we have placed on $\gamma_1 l'$ and $\gamma_2 l'$, we have

$$FN_s = \frac{Z^2_{13}}{4Z(1-\eta_1\eta_2)(K_1+\eta_1^2K_2)} \times \left[\frac{\eta_1}{K_1(\gamma_1-\gamma)}+\frac{\eta_2}{K_2(\gamma_2-\gamma)}\right]\left[\frac{\eta_1K_2}{\gamma_1+\gamma}+\frac{\eta_2K_1}{\gamma_2+\gamma}\right] \quad (62a)$$

$$= \frac{X\xi Z_{11}}{2(1-\eta_1\eta_2)(K_1+\eta_1^2K_2)} \times \left[\frac{\eta_1}{K_1(\gamma_1-\gamma)}+\frac{\eta_2}{K_2(\gamma_2-\gamma)}\right]\left[\frac{\eta_1K_2}{\gamma_1+\gamma}+\frac{\eta_2K_1}{\gamma_2+\gamma}\right]. \quad (62b)$$

Near-End

Under the above restriction that $e^{\gamma_1 l}$, $e^{\gamma_1 l'}$, $e^{\gamma_2 l}$ and $e^{\gamma_2 l'}$ are large compared with $e^{\gamma l}$, the currents and voltages in the disturbing section near $x = 0$ are given by equations (31) to (34) with the b-terms omitted, and the currents and voltages in the disturbed section adjacent to the sending end by equations (50) to (53).

The boundary conditions to be satisfied are that at $x = x' = 0$, $V_1 = V_1' = 0$ and $I_2 = -I_2'$. From these boundary conditions we obtain

$$a_1 = -(p_1+q_1) + \frac{\eta_1K_2(q_2+\eta_1q_1)}{K_1+\eta_1^2K_2}, \quad (63)$$

$$a_2 = -(p_2+q_2) + \frac{K_1(q_2+\eta_1q_1)}{K_1+\eta_1^2K_2}, \quad (64)$$

$$a_1' = \frac{\eta_1K_2(q_2+\eta_1q_1)}{K_1+\eta_1^2K_2}, \quad (65)$$

$$a_2' = \frac{K_1(q_2+\eta_1q_1)}{K_1+\eta_1^2K_2}. \quad (66)$$

The near-end near-end interaction crosstalk NN_s is given by

$$NN_s = \frac{Z_{13}}{2Z}\int_0^{l'} I_1' e^{-\gamma x} dx \tag{67a}$$

$$= -\frac{Z_{13}^2}{4Z(1-\eta_1\eta_2)(K_1+\eta_1^2K_2)} \times \left[\frac{\eta_1}{K_1(\gamma_1+\gamma)} + \frac{\eta_2}{K_2(\gamma_2+\gamma)}\right]\left[\frac{\eta_1 K_2}{\gamma_1+\gamma} + \frac{\eta_2 K_1}{\gamma_2+\gamma}\right] \tag{67b}$$

$$= -\frac{X\xi Z_{11}}{2(1-\eta_1\eta_2)(K_1+\eta_1^2K_2)} \times \left[\frac{\eta_1}{K_1(\gamma_1+\gamma)} + \frac{\eta_2}{K_2(\gamma_2+\gamma)}\right]\left[\frac{\eta_1 K_2}{\gamma_1+\gamma} + \frac{\eta_2 K_1}{\gamma_2+\gamma}\right]. \tag{67c}$$

Near-End Crosstalk with One Tertiary Short-Circuited

Although the derivation of the formula for near-end crosstalk N_s with one tertiary short-circuited is too long to be included here, it seems advisable to give this formula without derivation. Under the above mentioned restriction that $e^{\gamma_1 l}$ and $e^{\gamma_2 l}$ are large compared with $e^{\gamma l}$,

$$N_s = N_t + NN - NN_s, \tag{68}$$

where

N_t = near-end crosstalk, tertiaries terminated,

NN = near-end near-end interaction crosstalk between two adjoining lengths, tertiaries with no discontinuity,

NN_s = near-end near-end interaction crosstalk between two adjoining lengths with tertiary 1 short-circuited at the junction.

The first two terms (N_t and NN) may, with the types of cable studied so far, be determined with satisfactory accuracy from the single-tertiary analysis. In such a case, the formulas given herein are sufficient for computing the near-end crosstalk with one tertiary short-circuited.

III—Comparison of Computed Crosstalk with Measured Values

With 72-ft. and 145-ft. samples of the twin coaxial cable described in the companion paper by Messrs. Booth and Odarenko, crosstalk and impedance measurements were made in the laboratory, at frequencies from 50 kc to 300 kc, the sheath and quads in parallel being considered as providing a single tertiary, that is, as being connected together at short intervals.

The far-end crosstalk for a length of 5 miles was computed from these laboratory measurements and in Fig. 1 the results are compared with measurements on this length, the crosstalk in either case being practically the same whether the tertiary was terminated or short-circuited.

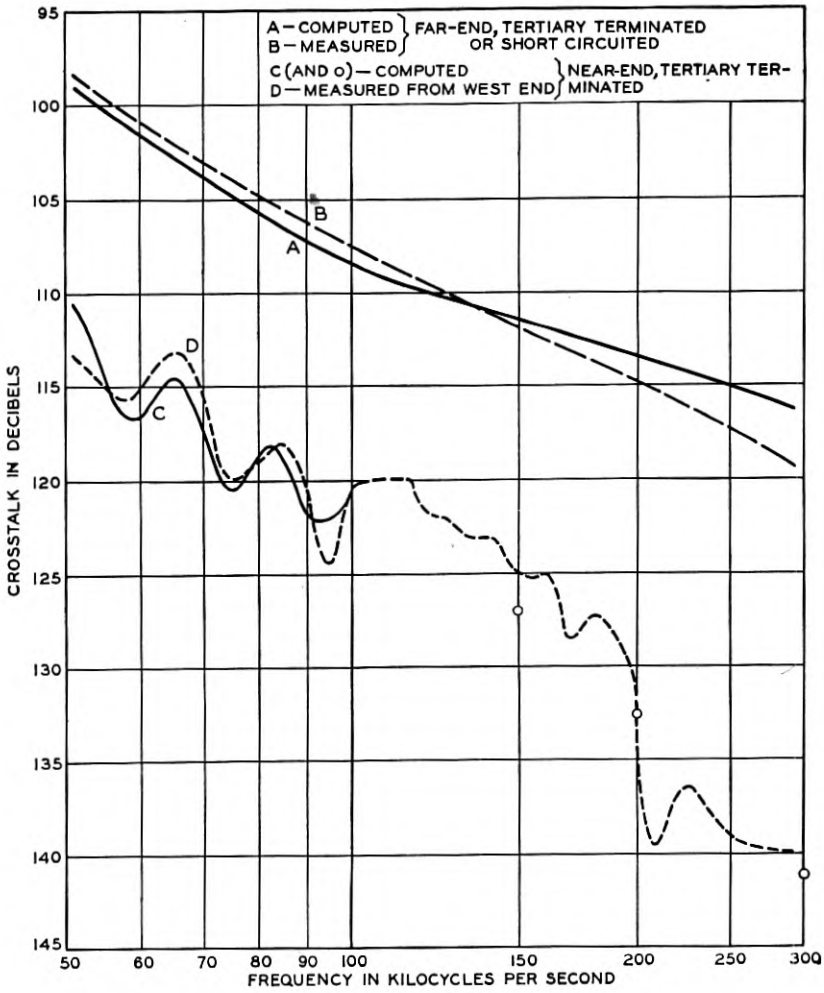

Fig. 1—Far-end crosstalk and near-end crosstalk for 5-mile length.

In Figs. 1 and 2, the computed near-end crosstalk for a length of 5 miles is compared with representative measurements on the above mentioned twin coaxial cable. Figure 1 shows this comparison with the tertiary terminated and Fig. 2 with the tertiary short-circuited.

The assumption of uniformity of the coaxial lines, as regards transfer impedances, and of the tertiary circuits as regards transmission characteristics, is a more serious restriction in the computation of near-end crosstalk than in the case of far-end crosstalk. Even for long lengths of cable, the near-end crosstalk is determined almost entirely by the crosstalk behavior of a relatively short length of the cable near the sending end, whereas the average crosstalk characteristics determine

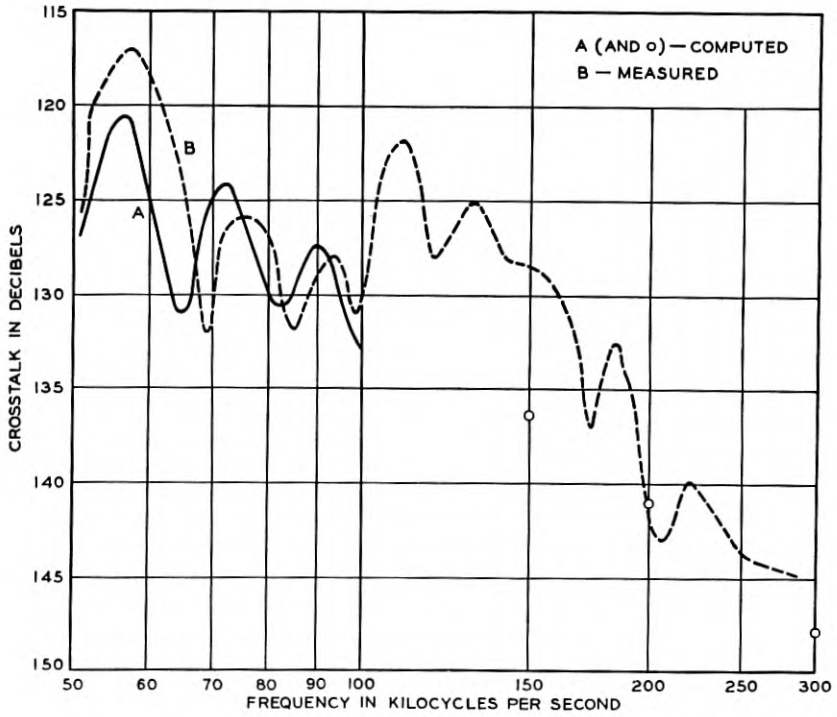

Fig. 2—Near-end crosstalk for 5-mile length, tertiary short-circuited.

the far-end crosstalk for a long length of cable. Thus, from measurements on representative short lengths, the far-end crosstalk for a long length may generally be computed more accurately than can the near-end crosstalk.

Similarly, the various types of interaction crosstalk depend largely upon the crosstalk behavior of relatively short lengths of the cable near the junction. In Fig. 3, the far-end far-end interaction crosstalk has been chosen as an illustration of the correlation which has been obtained between computed interaction crosstalk for the above men-

tioned twin coaxial cable and the measured interaction crosstalk The curves in Fig. 3 are for equal lengths either of 3000 ft. or 12,000 ft. In the case of the measured values, the junction point of the two sections was not the same for these two lengths. Although the agreement between calculated and measured values is only fair, the spread in the experimental results for these two cases, which for uniform cable would be slight, is about the same as the spread between calculated and measured values.

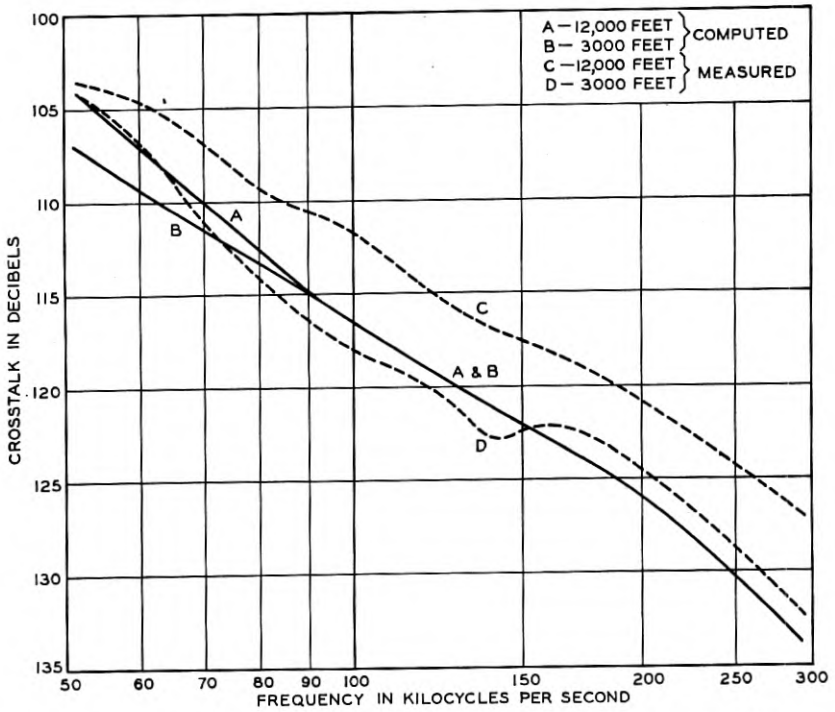

Fig. 3—Far-end far-end interaction crosstalk between two equal lengths.

The two-tertiary formulas have so far been applied only to one type of cable with four coaxial lines and a layer of paper-insulated pairs. The longest length of this type of cable on which crosstalk measurements have been made is 1900 ft. The various types of interaction crosstalk with one tertiary short-circuited, as computed from the formulas given above, agree roughly with the measured interaction crosstalk under this same condition. However, the restriction that the tertiary circuits involved are electrically long, as postulated in deriving the interaction crosstalk formulas for this case, is not satisfied,

and comparisons of the calculated and measured values are not very significant.

It may be remarked that the application of the single-tertiary analysis to all cases in which the two tertiaries were treated alike (either terminated or short-circuited) gave very satisfactory agreement between computed and measured crosstalk for this 1900 ft. length of 4-coaxial cable.

Crosstalk Between Coaxial Conductors in Cable

By R. P. BOOTH and T. M. ODARENKO

The available literature on crosstalk between coaxial conductors in contact makes it clear that the presence of any other conducting material in continuous or frequent contact with the coaxial outer conductors simply reduces the coupling per unit length without altering the law of crosstalk summation with length.

When the conducting material is insulated from the coaxials, as in the case of quads and sheath in coaxial cables, the situation is more complicated. Instead of simply reducing the coupling per unit length the quads and sheath, with the outer conductors for a return, provide a tertiary circuit in which interaction crosstalk can take place between elementary line sections. The summation with length for this type of crosstalk is quite different from that between two coaxials in contact and therefore the combined summation is obviously more involved.

Tests on sections of a five-mile length of coaxial cable were made at Princeton, New Jersey, in the latter part of 1937 and early in 1938 in order to obtain experimental verification of the manner in which the quads and sheath affect crosstalk summation with length. It is shown that the crosstalk component due to the presence of the sheath and quads opposes the component which is present between two coaxials in free space so that the resultant crosstalk is considerably lower than would be computed ignoring the tertiary effects.

Introduction

In spite of the geometrical and electrical symmetry of the coaxial circuit and the excellent shielding properties of the outer conductor, a part of the electromagnetic energy escapes from the circuit through the outer conductor and sets up an electromagnetic field in the space around it. Any circuit, be it even another coaxial placed in this field will absorb a part of the energy stored in the field and deliver it to the terminals of the circuit in the form of an unwanted or interfering current—the crosstalk current. The magnitude of this crosstalk current depends on a variety of factors, such as the physical characteristics of the conductors and of the intervening space, the frequency and the length of the circuit.

Expressions for two important cases of crosstalk between two coaxial circuits in *free space*, namely, the so-called "direct" crosstalk with the outer conductors in continuous contact and the "indirect" crosstalk with the outer conductors insulated from each other, were determined

and discussed in a previously published paper.[1] It was shown there that the direct far-end crosstalk is directly proportional to l and the direct near-end crosstalk is proportional to

$$\frac{1 - \epsilon^{-2\gamma l}}{2\gamma},$$

where l is the length and γ is the propagation constant of either coaxial unit. The indirect crosstalk was shown to be a more complicated function of the length.

The present paper extends this earlier work to include the case where the coaxials are enclosed in a common sheath or, in the general case, paralleled by any conducting material symmetrically disposed.[2] When this conducting material is introduced in the neighborhood of two coaxials in contact the conditions for crosstalk production are naturally changed from those existing in free space. If the material is uniformly distributed along the coaxials and is in continuous or frequent contact with the outer conductors the summation of crosstalk with length is the same as before but the magnitude is reduced. This reduction is due to the fact that part of the current formerly flowing on the disturbed outer conductor now flows on the new conducting material instead, thus reducing the direct crosstalk coupling per unit length.

In most cables, the coaxial outer conductors are in contact but the other conducting material (sheath and quads) is insulated from the outer conductors. The quads must obviously be insulated for normal use and the sheath is kept insulated except at the ends of a repeater section in order to permit the use of insulating joints for electrolysis prevention where required. This material thus provides an extra transmission circuit, or tertiary circuit, in which tertiary currents can be propagated up and down the line. In such a case the resulting crosstalk in any length consists of both the direct crosstalk between the contacting coaxials and the indirect crosstalk via the outer conductor-sheath and quad tertiary circuit. The general formulas given in the Schelkunoff-Odarenko paper apply for these components. Since the two components follow different laws regarding summation with length the resultant summation is quite complicated except for very short or very long lengths.

The study of the tertiary effects on crosstalk summation is the main contribution of this paper to crosstalk theory. Emphasis will be placed on the development of a simple physical picture which will help one to

[1] Schelkunoff-Odarenko paper in *Bell Sys. Tech. Jour.*, April, 1937.

[2] In the interim between our tests and this publication a paper by H. Kaden concerning this general subject was published in the Europaischer Fernsprechdienst, no. 50, October, 1938, pp. 366–373.

visualize clearly the influence of the tertiary circuits in the summation process. To produce such a picture a certain amount of review of the general crosstalk problem will be necessary. This is undertaken in Part I of this paper.

Part II is devoted mainly to the presentation of test data taken in November and December, 1937, January and February, 1938 on sections of a five-mile length of a twin coaxial cable near Princeton. These data confirm and graphically illustrate certain relationships developed in Part I. In addition they provide information on the tendency of tertiary circuits to complicate the effectiveness of transpositions and show how interaction crosstalk takes place around repeaters via the tertiary circuits.

PART I—THEORY

In any series of crosstalk tests on short lengths of paired or quadded cable where the problem of combining a number of such lengths is concerned it has generally been the practice to terminate both the test circuits and important tertiary circuits in characteristic impedance. Under such a condition the normal influence of all circuits in the production of crosstalk within each short section is provided for and the summation process, including interaction between successive sections, can be studied under actual line conditions. This is a general method applicable to any type of coupling and was adopted for the Princeton investigation. The effect of discontinuities such as short-circuited tertiaries at the extreme ends of a repeater section can be readily handled mathematically as correction terms due to "end effect."

To simplify the presentation of the factors involved, the discussion in this section will be confined mainly to the case of far-end crosstalk. In a twin coaxial cable where the transmission in the two units is in opposite directions there actually exists no far-end crosstalk problem since only talker echo, a near-end crosstalk phenomenon, is involved.[3] In multi-unit cable, however, there will be far-end crosstalk between different systems. Since this type of crosstalk tends to increase directly with the number of repeater sections it is important to understand its nature thoroughly. Moreover, in a study of fundamentals it is possible to avoid certain complications not essential to an understanding of the problem by investigating far-end rather than near-end crosstalk.

To present a clear picture of the physical meaning of some of the forthcoming mathematical expressions their derivations will be ap-

[3] This statement may not hold if the repeater impedances fail to match the line impedance since in that case the far-end crosstalk can be reflected and appear as near-end crosstalk.

proached in as elementary a fashion as possible. In order to do this we shall start with the simple arrangement of two coaxial conductors in free space, a case already covered in previous papers. To the crosstalk equations covering this case will then be added terms to allow for the effects of quads and sheath. In all that follows in Part I the quads and sheath will be considered as one unit referred to as the "sheath." This is a good approximation as will be shown in Part II.

The conception of two independent crosstalk components—a direct or transverse component between coaxials in contact and an indirect or interaction component via the sheath tertiary circuit—is not necessary for the solution of the problem. It is preserved here, however, as offering a familiar and much simpler approach to a clear understanding of the processes involved in crosstalk summation with length.

Far-End Crosstalk

Consider first an elementary section, dl, of a long single coaxial in free space as indicated in Sketch (a) of Fig. 1. If the current at this point in the center conductor is I_1 the current in the outer conductor is practically $-I_1$ since there is no other return path (except through the air dielectric which offers a high impedance especially at the lower broad-band frequencies considered here). Using Schelkunoff's nomenclature we may state that an open-circuit voltage equal to $e_1 = I_1 Z_{\alpha\beta} dl$ is developed on the outer surface of the outer coaxial conductor. The term $Z_{\alpha\beta}$ represents the surface transfer impedance (mutual impedance) per unit length between the inner and outer surfaces of the outer coaxial conductor.

Now suppose that we place another long coaxial parallel to the first one and, for generality, insulated from it as shown by Sketch (b) of Fig. 1. The open-circuit voltage e_1 on length dl of the first coaxial outer conductor will now cause current to flow in the intermediate circuit composed of the two outer conductors. The parameters of this circuit are γ_3 and Z_3 as shown on the sketch. In returning on the second coaxial outer conductor this current causes crosstalk into the second coaxial circuit.

It is convenient at this point to replace the original impressed voltage e_1 by the set of emf's shown in Sketch (c) of Fig. 1. The insertion of equal and opposite voltages $e_1/2$ on the outer surface of the disturbed coaxial outer conductor does not change conditions but enables us to consider certain effects separately. The first effect to be considered is that due to the pair of equal and opposite voltages $e_1/2$ in the loop composed of the two coaxial outer conductors. These voltages combine to form a "balanced" voltage e_1 which tends to drive current

around the balanced circuit composed of the two outer conductors. For the present we shall not consider the voltages $e_1/2$ which are in the same direction in the outer conductors.

The current in the "balanced" intermediate circuit of characteristic impedance Z_3 and propagation constant γ_3 due to the balanced voltage e_1 in the elementary length dl is $i_3 = e_1/2Z_3$. This current flowing along the outer coaxial conductor of the disturbed circuit produces a voltage $e_2 = i_3 Z_{\alpha\beta} dl$ on the inner surface of this outer conductor and this voltage in turn causes a current i_{2a} in the disturbed coaxial circuit

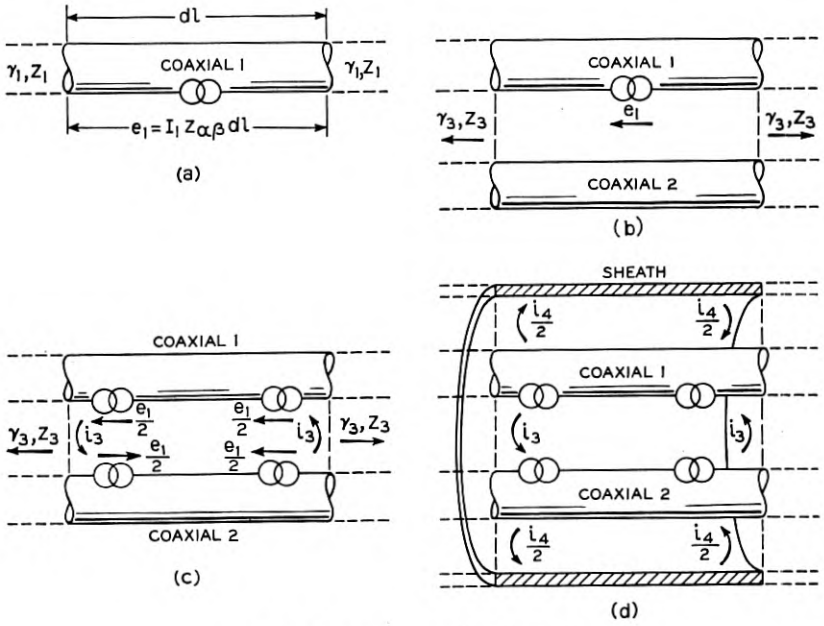

Fig. 1.—Coaxial crosstalk schematics.

equal to $e_2/2Z$, where Z is the coaxial characteristic impedance.[4] In a long line other elementary lengths of the disturbed coaxial are also affected by i_3 because of its propagation along the intermediate circuit. (This crosstalk by way of a tertiary circuit from one length into another is known as indirect or "interaction crosstalk" and because of its presence the summation of crosstalk with length is not a simple function of length even for systematic coupling such as occurs with coaxials.) This is a crosstalk case for which the general solution is already

[4] The subscript "a" in i_{2a} relates this current to the so-called "mode a" current used by Carson and Hoyt in their paper entitled "Propagation of Periodic Currents Over a System of Parallel Wires," *Bell Sys. Tech. Jour.*, July, 1927.

available. When the effects are integrated it is found that the far-end crosstalk is quite a complicated function of length and of the tertiary and coaxial propagation and impedance characteristics.[5] However, if the coaxial units are in actual contact as in the case of the coaxial cable to be considered here, the formula for the far-end crosstalk F_3 expressed as a current ratio is quite simple, namely,

$$F_3 = \frac{Z_{\alpha\beta}^2}{2ZZ_{33}} \cdot l, \tag{1}$$

where $Z_{33} = Z_3\gamma_3$ is the series impedance per unit length of the circuit composed of one coaxial outer conductor with return on the other. Thus, for this component, the far-end crosstalk is directly proportional to length. This simple relation results from the fact that the intermediate circuit, being continuously shorted, has such high attenuation that no interaction crosstalk between elementary lengths can exist.

We shall now consider the crosstalk contribution due to the longitudinal voltage $e_1/2$ acting along both coaxial outer conductors in parallel. Suppose that a sheath is placed symmetrically around the two coaxials but insulated from them as shown in Sketch (d) of Fig. 1. The longitudinal voltage sends a current around the circuit composed of the two parallel outer conductors with sheath return equal to $i_4 = e_1/(2)(2Z_4)$, where Z_4 is the characteristic impedance of this circuit. Half of this longitudinal current flows on the disturbed coaxial outer conductor *in opposition* to the balanced current i_3 flowing there.

Following previous procedure it can be shown that in the elementary length a crosstalk current $i_{2c} = i_4 Z_{\alpha\beta} dl/4Z$ will flow in the disturbed coaxial circuit.[6] Other elementary lengths are also affected by i_4 thus producing interaction crosstalk. When the effects are integrated over a length l the far-end crosstalk for this component is found to be as follows:

$$F_4 = -\frac{Z_{\alpha\beta}^2}{16ZZ_4}\left[\frac{2l}{\gamma_4}\cdot\frac{\gamma_4^2}{\gamma_4^2-\gamma^2} - \left(\frac{2(\gamma_4^2+\gamma^2)}{(\gamma_4^2-\gamma^2)^2} - \frac{\epsilon^{-(\gamma_4-\gamma)l}}{(\gamma_4-\gamma)^2} - \frac{\epsilon^{-(\gamma_4+\gamma)l}}{(\gamma_4+\gamma)^2}\right)\right], \tag{2}$$

where γ_4 is the propagation constant of the sheath-outer conductor circuit. If the sheath is in actual contact with the coaxial units the

[5] See equation (40) in the Schelkunoff-Odarenko paper in *Bell Sys. Tech. Jour.*, April, 1937.

[6] The subscript "c" in i_{2c} relates this current to the "mode c" current used by Carson and Hoyt in their paper of July, 1927.

formula reduces to the simple relation

$$F_4 = -\frac{Z_{\alpha\beta}^2}{16ZZ_4}\cdot\frac{2l}{\gamma_4} = -\frac{Z_{\alpha\beta}^2}{8ZZ_{44}}\cdot l, \tag{3}$$

where $Z_{44} = Z_4\gamma_4$ = series impedance per unit length of the circuit composed of the outer conductors with sheath return.[7] For a sheath in contact this component is thus directly proportional to length since interaction crosstalk from one elementary length into another has been eliminated.

Now, while we are actually concerned with the insulated sheath as covered by (2) it is of considerable interest to study equations (1) and (3) at this point. The total far-end crosstalk when the outer conductors and sheath are in contact is the sum of the crosstalk components F_3 and F_4 as given in equations (1) and (3), or

$$F_3 + F_4 = F_l = \frac{Z_{\alpha\beta}^2}{2Z}\left[\frac{l}{Z_{33}} - \frac{l}{4Z_{44}}\right]. \tag{4}$$

This simple addition follows from the fact that the circuits for the two modes of propagation covered by equations (1) and (3) are mutually non-inductive because of symmetry so that there is no reaction between them. The recognition of this fact does away with the necessity of complicated mathematics which would otherwise have to be used in the general solution.[8]

In formula (4) the second term in the bracket represents the contribution of the tertiary circuit involving the sheath and is seen to be opposite in sign to the first term which represents the crosstalk which would exist in the absence of the sheath. The equation illustrates mathematically the previous statement that conducting material in contact with the coaxials acts to reduce the crosstalk. Since both components are directly proportional to length, the total is also directly proportional to length.

It is apparent in formula (4) that the crosstalk would be zero if the values of Z_{33} and $4Z_{44}$ were equal. In cables where steel tapes are used on the outer surface of the coaxials this condition is approached. For example, if we neglect external inductance and proximity effects, Z_{33} would be equal to twice the surface self-impedance of a single outer

[7] It should be noted here that it is not really necessary to postulate a separate sheath return in order to obtain expression (3) for F_4 due to the longitudinal voltage $e_1/2$, since the return *in continuous contact* with the outer conductors will actually tend to lose its identity. The device of introducing sheath return insulated from the outer conductors and then shorting it to the conductors serves only to simplify the concepts of Z_{44}, Z_4, γ_4.

[8] This principle of symmetry can be extended to the case of four coaxial units whether insulated or in contact.

conductor while Z_{44} would equal one-half of the surface self-impedance of a single outer conductor (neglecting the self-impedance of the lead sheath in comparison with the iron outer conductors). Thus, neglecting differences in external inductance and in proximity effects $1/Z_{33}$ would equal $1/4Z_{44}$ and the crosstalk would vanish. Actually, the observed reduction in crosstalk due to these opposing terms is about 32 db at 50 kilocycles in a 145-foot section of twin coaxial with quads and sheath shorted to the coaxials at the ends only. Physically this means that the current due to the voltage on the outer conductor surface of the disturbing coaxial flows mainly in the sheath and quads rather than on the high impedance surface of the disturbed coaxial.

Now let us consider the case where the sheath is insulated from the coaxial outer conductors. For this case equations (1) and (2) may also be added directly to give

$$F_l = \frac{Z_{\alpha\beta}^2}{2Z}\left[\left(\frac{l}{Z_{33}} - \frac{l}{4Z_{44}}\cdot\frac{\gamma_4^2}{\gamma_4^2-\gamma^2}\right) + \frac{\gamma_4}{4Z_{44}}\left(\frac{\gamma_4^2+\gamma^2}{(\gamma_4^2-\gamma^2)^2}\right) - \frac{\gamma_4}{4Z_{44}}\left(\frac{\epsilon^{-(\gamma_4-\gamma)l}}{2(\gamma_4-\gamma)^2} + \frac{\epsilon^{-(\gamma_4+\gamma)l}}{2(\gamma_4+\gamma)^2}\right)\right]. \quad (5)$$

This equation appears quite formidable but it has been split purposely into three terms which will be examined individually. The first term is directly proportional to length, the second term is independent of length and the third term involves length exponentially. For lengths where the tertiary circuit is electrically long the third term vanishes and we have

$$F_l = \frac{Z_{\alpha\beta}^2}{2Z}\left[\left(\frac{l}{Z_{33}} - \frac{l}{4Z_{44}}\cdot\frac{\gamma_4^2}{\gamma_4^2-\gamma^2}\right) + \frac{\gamma_4}{4Z_{44}}\left(\frac{\gamma_4^2+\gamma^2}{(\gamma_4^2-\gamma^2)^2}\right)\right]. \quad (6)$$

In electrically short lengths we get

$$F_l = \frac{Z_{\alpha\beta}^2}{2Z}\left[\left(\frac{l}{Z_{33}} - \frac{l}{4Z_{44}}\cdot\frac{\gamma_4^2}{\gamma_4^2-\gamma^2}\right) + \frac{\gamma_4}{4Z_{44}}\left(\frac{\gamma_4 l}{\gamma_4^2-\gamma^2} - \frac{l^2}{2}\right)\right] = \frac{Z_{\alpha\beta}^2}{2Z}\left[\frac{l}{Z_{33}} - \frac{\gamma_4}{4Z_{44}}\cdot\frac{l^2}{2}\right] \backsimeq \frac{Z_{\alpha\beta}^2}{2Z}\left[\frac{l}{Z_{33}}\right], \quad (7)$$

in which it is seen that terms two and three of (5) combine to cancel the second half of term one.

From equations (5), (6) and (7) we are now ready to build a physical picture of what takes place as l is increased for cable sections where the sheath is insulated from the coaxial outer conductors but terminated to them at each end in characteristic impedance, Z_4. Starting with equation (7) we see that for very short lengths the term involving

l^2 becomes negligible, that is, the crosstalk is practically all due to the component which exists in the complete absence of a sheath (see equation (1)). In the range of lengths where this is true *the crosstalk increases directly with length.*

Quite a different state of affairs exists for a section electrically long enough for equation (6) to hold. The first bracketed term is still proportional to length but now consists of the difference of two components. The first of these represents the crosstalk between the coaxials *with no sheath present* while the second is a part of the crosstalk component introduced by the presence of the sheath. Except for the factor $\gamma_4^2/\gamma_4^2 - \gamma^2$ this first bracketed term in equation (6) is the same as equation (4) for a sheath in contact where, as we have already noted, the cancellation of the two components is quite effective when steel tapes are used on the outer conductors. Since γ_4^2 is necessarily considerably greater than γ^2 because of these steel outer conductors, it is reasonable to expect that the factor $\gamma_4^2/\gamma_4^2 - \gamma^2$ is nearly unity and that, therefore, the two components in the first bracketed term of equation (6) will also tend to cancel leaving a residual proportional to length but *much* lower in magnitude than either component *alone.*

The second bracketed term of equation (6) is entirely independent of length. This term has also been introduced by the presence of the tertiary circuit and its magnitude depends on the characteristics of this circuit.

Thus, even without knowing the relative magnitudes of the two components of the first bracketed term of equation (6) for a given length, it is apparent that as l is increased this term must eventually be controlling. The crosstalk will then again be proportional to length as it was for very short lengths but at a reduced level proportional to

$$\frac{\dfrac{1}{Z_{33}} - \dfrac{1}{4Z_{44}} \cdot \dfrac{\gamma_4^2}{\gamma_4^2 - \gamma^2}}{\dfrac{1}{Z_{33}}} = 1 - \frac{Z_{33}}{4Z_{44}} \cdot \frac{\gamma_4^2}{\gamma_4^2 - \gamma^2}.$$

It is quite evident, too, that for a range of lengths where the tertiary circuits are electrically long but where the first term of equation (6) has not had a chance to build up sufficiently the crosstalk will be about constant at a level determined mainly by the second term.

The above analysis may well suffice as a background for an interpretation of the measurements to be given in Part II. However, another and perhaps in some ways a more illuminating approach from a physical standpoint is possible.

Suppose, for example, that far-end crosstalk measurements are made on two cable sections each of length l with tertiaries terminated as illustrated in Sketch (a) of Fig. 2. Let the total crosstalk in each section be equal to F_l as defined by equation (5) above. If these two sections are joined together the total crosstalk is $2F_l$ plus some other terms which represent the interaction crosstalk between the two sections as illustrated in Sketch (b) of Fig. 2. We shall call the component F_{nn}

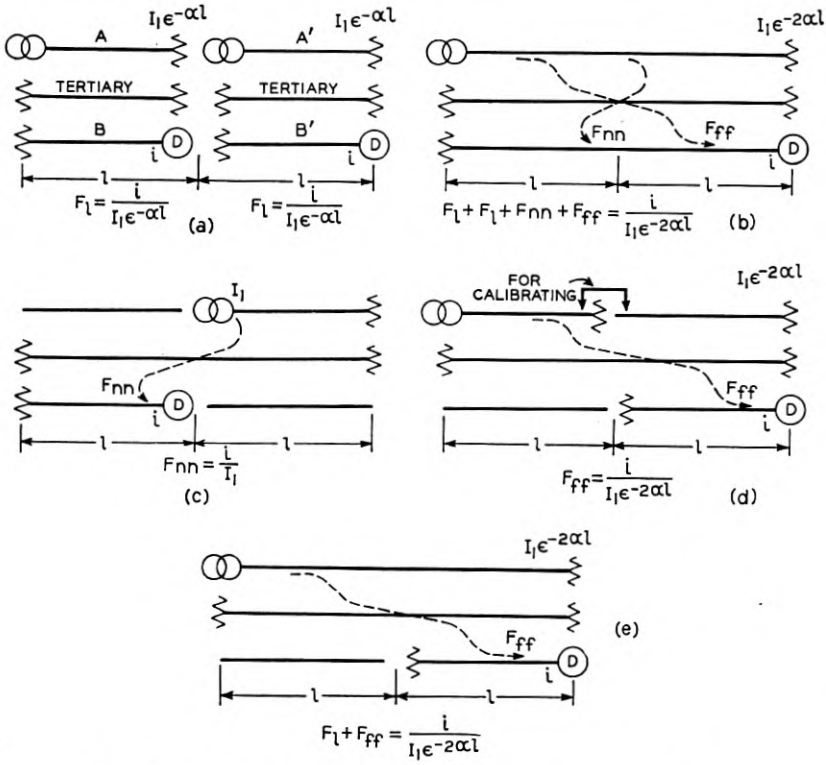

Fig. 2—Schematics illustrating far-end crosstalk summation.

near-end near-end and the component F_{ff} far-end far-end interaction crosstalk. Although inseparable under normal line conditions, these components are definite physical entities and can be isolated as shown schematically on Sketches (c) and (d) of Fig. 2. Thus, both F_{nn} and F_{ff} can be measured readily. In addition it is possible to measure directly $F_l + F_{ff}$ as shown on Sketch (e).

This interaction crosstalk between sections is due to crosstalk currents introduced into the outer conductor-sheath tertiary circuit in one

section and propagated along this circuit into the next section and thence into the disturbed coaxial. Except for interaction crosstalk between sections the total crosstalk in $2l$ would simply be twice that in length l, that is, the crosstalk would be directly proportional to length.

Now, the expressions for far-end crosstalk due to such interactions between two sections each of length l are

$$F_{nn} = -\frac{Z_{\alpha\beta}^2}{4Z} \cdot \frac{\gamma_4}{4Z_{44}} \left[\frac{1 - \epsilon^{-(\gamma_4+\gamma)l}}{\gamma_4 + \gamma} \right]^2, \tag{8}$$

$$F_{ff} = -\frac{Z_{\alpha\beta}^2}{4Z} \cdot \frac{\gamma_4}{4Z_{44}} \left[\frac{1 - \epsilon^{-(\gamma_4-\gamma)l}}{\gamma_4 - \gamma} \right]^2. \tag{9}$$

Since the coefficients [9] outside of the brackets are the same for F_{nn} and F_{ff} the terms may be combined to give the total interaction crosstalk between the two sections, namely,

$$F_{nn} + F_{ff} = -\frac{Z_{\alpha\beta}^2}{2Z} \left[\frac{\gamma_4}{4Z_{44}} \left(\frac{\gamma_4^2 + \gamma^2}{(\gamma_4^2 - \gamma^2)^2} \right) - \frac{\gamma_4}{4Z_{44}} \left(\frac{\epsilon^{-(\gamma_4-\gamma)l}}{(\gamma_4 - \gamma)^2} + \frac{\epsilon^{-(\gamma_4+\gamma)l}}{(\gamma_4 + \gamma)^2} \right) + \frac{\gamma_4}{4Z_{44}} \left(\frac{\epsilon^{-2(\gamma_4-\gamma)l}}{2(\gamma_4 - \gamma)^2} + \frac{\epsilon^{-2(\gamma_4+\gamma)l}}{2(\gamma_4 + \gamma)^2} \right) \right]. \tag{10}$$

As mentioned before, the crosstalk in length $2l$ *exclusive* of interactions *between* the two sections is equal to $2F_l$ or equation (5) multiplied by 2, namely,

$$2F_l = \frac{Z_{\alpha\beta}^2}{2Z} \left[\left(\frac{2l}{Z_{33}} - \frac{2l}{4Z_{44}} \cdot \frac{\gamma_4^2}{\gamma_4^2 - \gamma^2} \right) + \frac{2\gamma_4}{4Z_{44}} \left(\frac{\gamma_4^2 + \gamma^2}{(\gamma_4^2 - \gamma^2)^2} \right) - \frac{2\gamma_4}{4Z_{44}} \left(\frac{\epsilon^{-(\gamma_4-\gamma)l}}{2(\gamma_4 - \gamma)^2} + \frac{\epsilon^{-(\gamma_4+\gamma)l}}{2(\gamma_4 + \gamma)^2} \right) \right]. \tag{11}$$

The total crosstalk in length $2l$ is then the sum of (10) and (11), namely,

$$F_{2l} = 2F_l + F_{nn} + F_{ff} = \frac{Z_{\alpha\beta}^2}{2Z} \left[\left(\frac{2l}{Z_{33}} - \frac{2l}{4Z_{44}} \cdot \frac{\gamma_4^2}{\gamma_4^2 - \gamma^2} \right) + \frac{\gamma_4}{4Z_{44}} \left(\frac{\gamma_4^2 + \gamma^2}{(\gamma_4^2 - \gamma^2)^2} \right) - \frac{\gamma_4}{4Z_{44}} \left(\frac{\epsilon^{-2(\gamma_4-\gamma)l}}{2(\gamma_4 - \gamma)^2} + \frac{\epsilon^{-2(\gamma_4+\gamma)l}}{2(\gamma_4 + \gamma)^2} \right) \right], \tag{12}$$

[9] These near-end near-end and far-end far-end coefficients are equal because the coupling through a coaxial is of a series voltage character. In open wire and non-shielded cables where there is also present coupling due to shunt admittances the coefficients for F_{nn} and F_{ff} are different in magnitude and their effects must be considered separately. See paper by A. G. Chapman in *Bell Sys. Tech. Jour.* for January and April, 1934.

wherein the second term in equation (10) is cancelled completely by the third term of equation (11). This equation (12) is exactly what we would get by substituting $2l$ for l in the general equation (5). The only reason for deriving it in terms of $2F_l$ plus interaction between the sections is to present a better physical picture of the mechanism of far-end crosstalk summation with length, that is, to show how the interaction crosstalk *between* two sections alters what otherwise would be a direct summation with length.

In lengths where the tertiary circuit is electrically long equation (12) for total crosstalk in length $2l$ becomes

$$F_{2l} = 2F_l + F_{nn} + F_{ff} = \frac{Z_{\alpha\beta}^2}{2Z}\left[\left(\frac{2l}{Z_{33}} - \frac{2l}{4Z_{44}}\cdot\frac{\gamma_4^2}{\gamma_4^2 - \gamma^2}\right) + \frac{\gamma_4}{4Z_{44}}\left(\frac{\gamma_4^2 + \gamma^2}{(\gamma_4^2 - \gamma^2)^2}\right)\right], \quad (13)$$

which differs from equation (6) for total crosstalk in length l only by the factor of 2 in the first bracketed term. Thus, as mentioned before, there is a range of lengths wherein the crosstalk will be constant at a level determined by the second term of (6) or (12) until the length becomes sufficient for the first term to become controlling.

In lengths where the tertiary circuit is electrically short equation (11) becomes

$$2F_l = \frac{Z_{\alpha\beta}^2}{2Z}\left[\frac{2l}{Z_{33}} - \frac{\gamma_4}{4Z_{44}}\cdot l^2\right], \quad (14)$$

which reduces simply to

$$2F_l = \frac{Z_{\alpha\beta}^2}{2Z}\left[\frac{2l}{Z_{33}}\right] = \left[\frac{Z_{\alpha\beta}^2}{ZZ_{33}}\right] l \quad (15)$$

when the length is sufficiently short. The interaction crosstalk *between* two electrically short lengths becomes, from equation (10),

$$F_{nn} + F_{ff} = \frac{-Z_{\alpha\beta}^2}{2Z}\left[\frac{\gamma_4}{4Z_{44}}\cdot l^2\right] = \left[\frac{-Z_{\alpha\beta}^2}{8ZZ_4}\right] l^2, \quad (16)$$

one-half of which is due to component F_{nn} and the other half to component F_{ff}. The sum of (14) and (16) is

$$F_{2l} = 2F_l + F_{nn} + F_{ff} = \frac{Z_{\alpha\beta}^2}{2Z}\left[\frac{2l}{Z_{33}} - \frac{\gamma_4}{4Z_{44}}\cdot 2l^2\right], \quad (17)$$

which is exactly equal to equation (7) if $2l$ is substituted for l therein. From (15) and (16) it is apparent that for very short lengths the total

crosstalk in length $2l$ will be simply twice that in length l since the interaction crosstalk between lengths l is proportional to l^2 and therefore is negligibly small.

The view of the mechanism of far-end crosstalk summation as developed above is illustrated by measurements to be presented in Part II. It may be pointed out here that the measurement of far-end and interaction crosstalk in phase and magnitude on short lengths where equations (15) and (16) hold gives the far-end and interaction crosstalk coefficients from which the crosstalk in any length of line may be computed provided the propagation constants and impedances of the coaxial and the tertiary circuits are known.

A practical difficulty may arise from the fact that the application of this method involves equations (12) or (5) where the first bracketed term consists of the difference of two quantities each of which is very large compared with this difference. Thus, a considerable error may be introduced in the computation of this term because of small errors in the measurement of its components. For some cases it is, therefore, better to use a method based on certain crosstalk measurements in a short length of cable with the tertiary circuits open and shorted.[10] There are cases, however, where the controlling crosstalk in a five-mile section is predominantly due to the second term of equation (5). One such case is for the crosstalk between diagonally opposite coaxials in a four-coaxial cable. In this case tests have shown that the cancellation of components in the first term is so complete that the second term is controlling in five miles. For such a case the more accurate method may be to determine the interaction coefficient from equation (16).

Near-End Crosstalk

It will be sufficient here to give simply the final equations for the two crosstalk components for any length l.

For the component which would exist for two contacting coaxials in free space we have

$$N_3 = \frac{Z_{\alpha\beta}^2}{2Z} \cdot \frac{1}{Z_{33}} \left(\frac{1 - \epsilon^{-2\gamma l}}{2\gamma} \right) \tag{18}$$

and for that component due to the presence of the sheath

$$N_4 = -\frac{Z_{\alpha\beta}^2}{2Z} \cdot \frac{1}{4Z_{44}} \left[\frac{\gamma_4^2}{\gamma_4^2 - \gamma^2} \cdot \frac{1 - \epsilon^{-2\gamma l}}{2\gamma} - \frac{\gamma_4^2}{\gamma_4^2 - \gamma^2} \cdot \frac{1 - 2\epsilon^{-(\gamma_4+\gamma)l} + \epsilon^{-2\gamma l}}{2\gamma_4} \right], \tag{19}$$

[10] The method described in a companion paper by K. E. Gould.

whence, for both components,

$$N_3 + N_4 = N_l = \frac{Z_{\alpha\beta}^2}{2Z}\left[\left(\frac{1}{Z_{33}} - \frac{1}{4Z_{44}}\cdot\frac{\gamma_4^2}{\gamma_4^2 - \gamma^2}\right)\frac{1 - \epsilon^{-2\gamma l}}{2\gamma} + \frac{1}{4Z_{44}}\cdot\frac{\gamma_4^2}{\gamma_4^2 - \gamma^2}\left(\frac{1 - 2\epsilon^{-(\gamma_4+\gamma)l} + \epsilon^{-2\gamma l}}{2\gamma_4}\right)\right]. \quad (20)$$

In a section where the tertiary circuit is electrically long equation (20) reduces to

$$N_l = \frac{Z_{\alpha\beta}^2}{2Z}\left[\left(\frac{1}{Z_{33}} - \frac{1}{4Z_{44}}\cdot\frac{\gamma_4^2}{\gamma_4^2 - \gamma^2}\right)\frac{1 - \epsilon^{-2\gamma l}}{2\gamma} + \frac{1}{4Z_{44}}\cdot\frac{\gamma_4^2}{\gamma_4^2 - \gamma^2}\left(\frac{1 + \epsilon^{-2\gamma l}}{2\gamma_4}\right)\right] \quad (21)$$

and when l is electrically short it reduces to

$$N_l = \frac{Z_{\alpha\beta}^2}{2Z}\left[\frac{l}{Z_{33}} - \frac{\gamma_4}{4Z_{44}}\cdot\frac{l^2}{2}\right], \quad (22)$$

which is the same as for far-end crosstalk in very short lengths as given in equation (7).

As pointed out earlier the expression for near-end crosstalk even when the tertiary circuit is electrically long is more complicated in form than for far-end crosstalk because of the terms $1 - \epsilon^{-2\gamma l}$ and $1 + \epsilon^{-2\gamma l}$. This may be seen by comparing formulas (6) and (21).

Nevertheless it is possible to see from (21) that the presence of the tertiary circuit acts to reduce near-end crosstalk as it did in the case of far-end crosstalk. The first term of (21) is less than the near-end crosstalk without the sheath (equation (18)) by the factor

$$\frac{\frac{1}{Z_{33}} - \frac{1}{4Z_{44}}\cdot\frac{\gamma_4^2}{\gamma_4^2 - \gamma^2}}{\frac{1}{Z_{33}}} = 1 - \frac{Z_{33}}{4Z_{44}}\cdot\frac{\gamma_4^2}{\gamma_4^2 - \gamma^2}.$$

This is the same factor by which far-end crosstalk is reduced in very long lengths as brought out in the discussion of equation (6). However, the second term in equation (21) prevents this complete reduction from ever taking place in the case of near-end crosstalk.

PART II—EXPERIMENTAL RESULTS

The crosstalk measurements presented here were made on and between sections of twin coaxial cable of various lengths from 73 feet

to about five miles. Primarily the tests were made to indicate the effect of sheath and quads upon the summation of crosstalk with length as a check on theoretical considerations and were the first extensive tests made on a coaxial cable with this end in view. The layup of the cable is shown in Fig. 3.

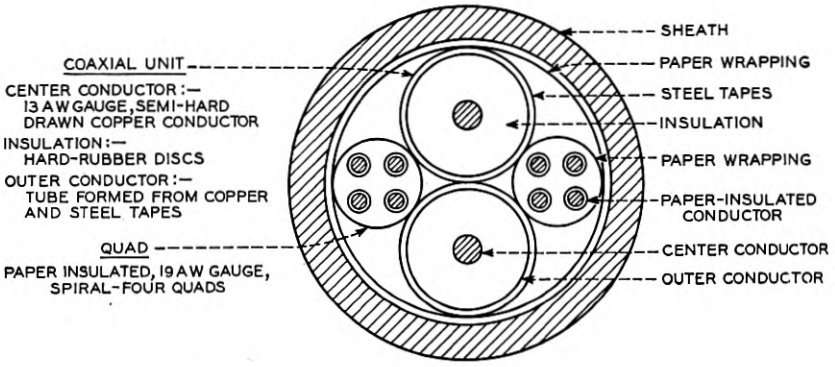

Fig. 3—Cross-section of twin coaxial cable.

As indicated in the latter portion of Part I the general procedure was to measure crosstalk in available sections of equal length, l, with the tertiary circuits terminated in approximately characteristic impedance. Interaction crosstalk between these sections was then measured and finally the two sections were combined to find the total crosstalk in length $2l$. This process was repeated until a total length of about five miles was built up.

Far-End Crosstalk Summation

The results of crosstalk tests on 73 and 146-foot lengths are shown in Fig. 4. The letters on the curves correspond to the crosstalk components discussed in Part I. Only far-end far-end interaction crosstalk was measured but for such short lengths the near-end near-end crosstalk would be nearly the same.

Remembering from the discussion in Part I that the total crosstalk F_{2l} in length $2l$ is equal to $2F_l + F_{nn} + F_{ff}$ it is evident that since in this case the measured components F_{nn} or F_{ff} are quite small the crosstalk in 146 feet should be approximately $2F_l$. That this is the case may be seen from the measured crosstalk in 146 feet which is about 6 db higher than for 73 feet. These lengths are apparently short enough for equations (15) and (16) to hold reasonably well at the lower frequencies.

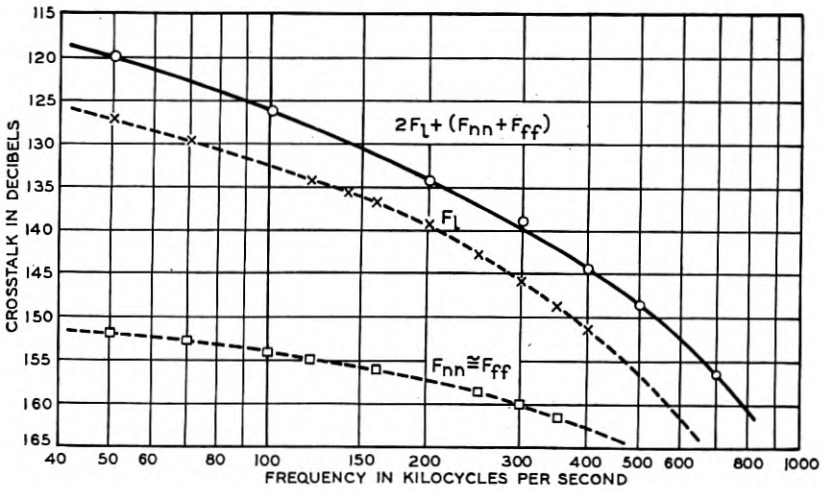

Fig. 4—Crosstalk components in 73-foot and 146-foot lengths.

Now suppose that we consider two lengths which are considerably longer so that equations (6) and (13) more nearly apply. Figure 5 shows the results of tests on and between two 1500-foot cable sections. Here, in contrast with the 73-foot measurements, components F_l and F_{ff} are nearly equal in magnitude while F_{nn} is quite small. Also, F_{ff} and F_l are in general phase opposition since their sum, $F_l + F_{ff}$, is considerably less than either component alone.

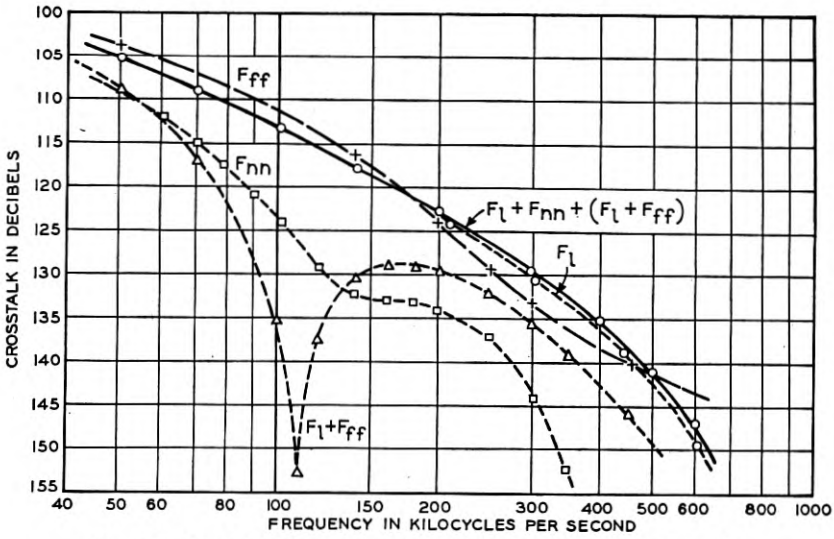

Fig. 5—Crosstalk components in 1500-foot and 3000-foot lengths.

Reference to equations (6) and (9) show that this tendency to cancel is to be expected provided the second term of (6) is the controlling term in F_l. Indeed, in lengths where the tertiary is electrically long, equations (8) plus (9) should exactly cancel the second term of (6). In other words, the total interaction crosstalk *between* two such sections should cancel a portion of the interaction crosstalk *within* a section. Since the portion which is cancelled is the controlling term the net result is that when two sections are combined the total crosstalk in length $2l$ is no more than was measured in length l, as evidenced by the measured curve $F_l + F_{nn} + (F_l + F_{ff})$ of Fig. 5.

This effect persists when two 3000-foot lengths are combined to form a 6000-foot section, as illustrated by the curves of Fig. 6. Here

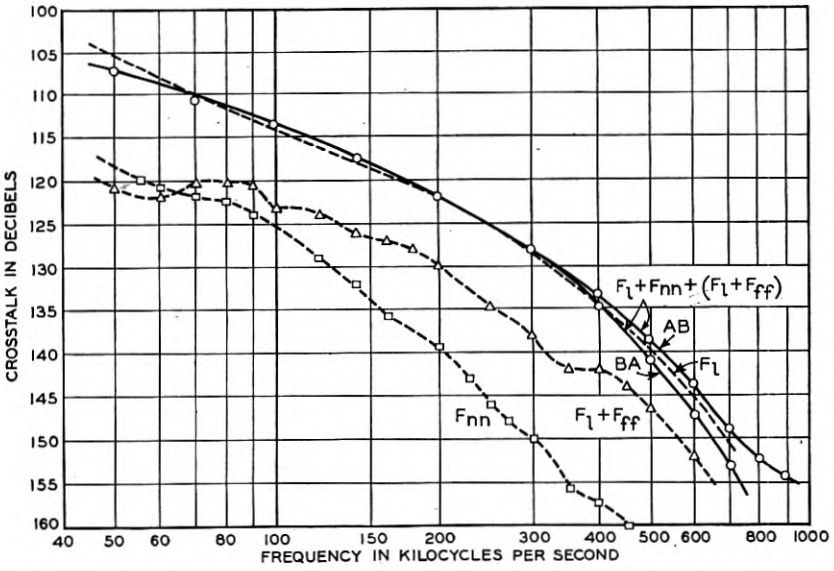

Fig. 6—Crosstalk components in 3000-foot and 6000-foot lengths.

again $(F_l + F_{ff})$ and F_{nn} are considerably smaller in magnitude than F_l so that the total crosstalk in 6000 feet cannot differ materially from the value F_l measured in 3000 feet.

The curves labelled AB and BA were made by using first coaxial A and then coaxial B as the disturbing circuit. The difference between the curves indicates that there is a certain amount of random unbalance within the section. For example, random deviations in the shielding of the two coaxials from a nominal value would result in different values of interaction crosstalk when the disturbed and dis-

turbing circuits are interchanged. The direct crosstalk component would not exhibit this effect.

The results of tests on 6000 and 12,000-foot lengths are given in Fig. 7. Again, the trend is in the same direction as in Figs. 5 and 6 except that in this case $(F_l + F_{ff})$ is nearly equal to F_l and has an appreciable influence when the two components are combined to give far-end crosstalk in 12,000 feet. This indicates that the first term of F_l in equation (6) is becoming more important as l is increased as would be expected.

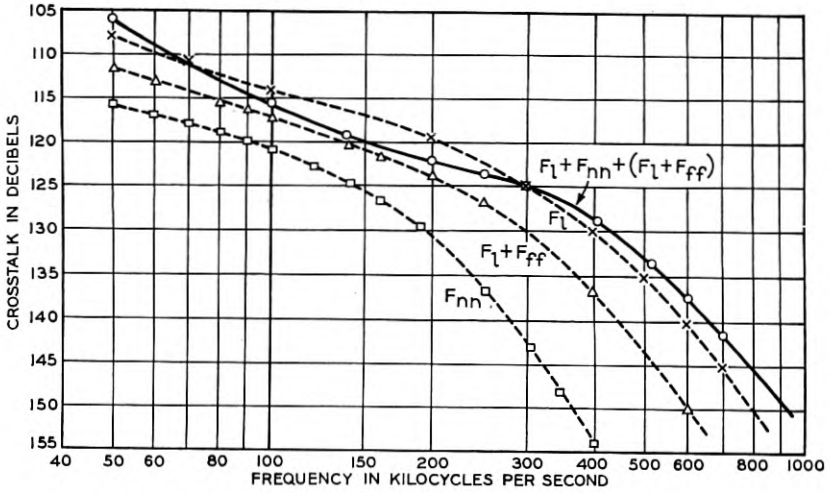

Fig. 7—Crosstalk components in 6000-foot and 12,000-foot lengths.

It may be noted here that curve F_l in Fig. 7 differs considerably from the AB and BA curves of $F_l + F_{nn} + (F_l + F_{ff})$ in Fig. 6 although all represent far-end crosstalk in 6000-foot sections. These differences in magnitude must be due to differences in the construction of the two cable sections. The difference between the curves varies from 3 to 8 db in the frequency range above 200 kilocyles. However, up to about 150 kilocycles the differences are not greater than 1 db. At the higher frequencies such differences naturally will introduce difficulties in any analysis since they superpose sizeable random effects on the major component of crosstalk which is systematic.

The curves in Fig. 8 present far-end crosstalk tests on 12,000 and 24,000-foot lengths. Here F_l and $(F_l + F_{ff})$ are of the same order of magnitude and combine in such a way that the crosstalk in 24,000 feet is from 3 to 6 db higher than that measured in 12,000 feet. Com-

ponent F_{nn} is again negligible. This behavior indicates that the first term of equation (6) is controlling as the length is increased.

It appears from all these tests that the magnitude of the far-end crosstalk in this cable with tertiaries terminated does not vary materially from 1500-foot to 12,000-foot cable lengths, except for random effects. In other words, for this range of lengths the second term of (6) is controlling. For very short lengths the crosstalk varies directly with length due to the absence of interaction crosstalk of sufficient magnitude to exert any influence. Also, in going from 12,000 to 24,000 feet, there is a definite indication that the crosstalk is increasing with

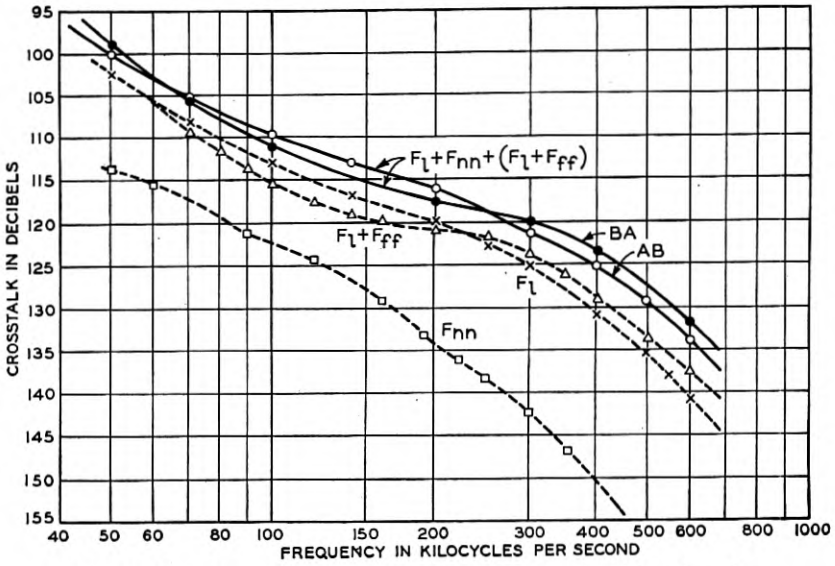

Fig. 8—Crosstalk components in 12,000-foot and 24,000-foot lengths.

length, so that for lengths over 24,000 feet the crosstalk would again tend to be proportional to length. We have shown in Part I that on the basis of theoretical considerations this law of crosstalk summation with length might be expected.

To illustrate this measured behavior the far-end crosstalk versus length for frequencies of 50, 100 and 200 kilocycles has been plotted on Fig. 9. For comparison are also plotted dashed curves based on the 73-foot tests and computed on the assumption that the crosstalk is directly proportional to length. The difference between corresponding curves shows the influence of the tertiary circuits. For a 24,000-foot length this difference amounts to 23, 26 and 27 db at 50, 100 and 200 kilocycles, respectively.

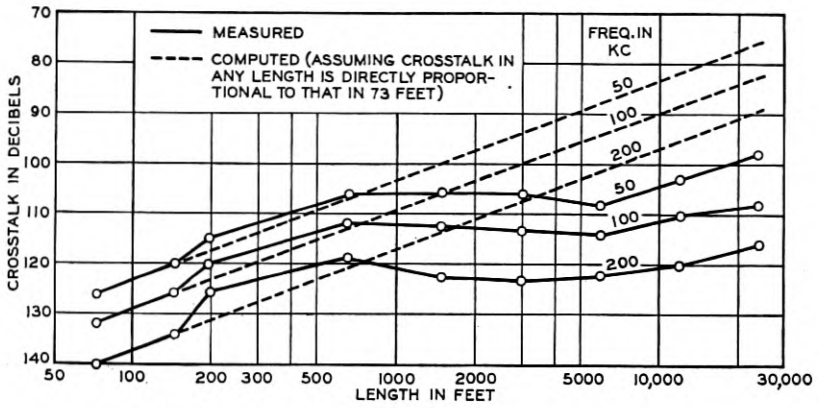

Fig. 9—Far-end crosstalk vs. length with tertiary terminated.

Near-End Crosstalk Summation

The curves on Fig. 10 show the amount of near-end crosstalk reduction due to the presence of the sheath and quads for a length of about five miles. The upper curve was computed from tests on a 73-foot length with tertiaries terminated by raising the values measured

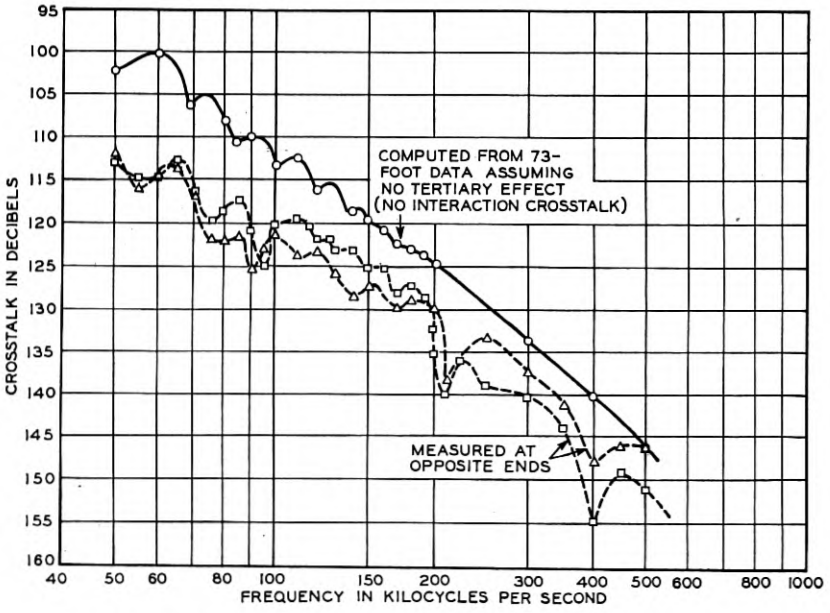

Fig. 10—Near-end crosstalk in a 5-mile length with tertiary terminated.

there by the factor

$$\frac{1 - \epsilon^{-2\gamma L}}{2\gamma l},$$

where $l = 73$ feet and $L = 5$ miles. This is the crosstalk which would exist in five miles *in the absence of a sheath and quads.*

The lower two curves were measured at opposite ends of the cable and the difference of about 10 db between these curves and the upper curve is due to the tertiary circuit effects. As might be expected from the discussion of equation (21), this reduction is considerably less than in the case of far-end crosstalk.

Interaction Crosstalk Between Sections

The methods of measuring the various types of interaction crosstalk between two sections have already been discussed in reference to Fig. 2. Besides showing the influence of interaction crosstalk in the summation of crosstalk *within* a repeater section the results presented below are indicative of the importance of interaction crosstalk which takes place *between* repeater sections, that is, around repeaters, when all or only a part of the tertiary is continuous at repeater points.

Values of near-end near-end interaction crosstalk, F_{nn}, were measured between various section lengths from 73 to 12,000 feet. It was found that the results are roughly independent of the section lengths above 1500 feet, and curve F_{nn} of Fig. 11 for the crosstalk measured between two 12,000 foot sections is typical. This independence of length is because of the high attenuation of the tertiary circuits which annihilates the effects of crosstalk in the more remote portions of the sections as may be seen from equation (8) if γ_4 is made large. The relatively unimportant contribution of this type of interaction crosstalk to the summation of far-end crosstalk *within* a repeater section has been discussed.

Similarly, measured values of far-end far-end and near-end far-end interaction crosstalk between various sections lengths were found to be practically independent of length above 1500 feet. Curves F_{ff} and N_{nf} of Fig. 11 for the crosstalk between 12,000-foot sections are typical. The far-end far-end component of interaction crosstalk has an important influence on the summation of far-end crosstalk within a repeater section as already mentioned in the section on far-end crosstalk summation. The influence of near-end far-end interaction crosstalk N_{nf}, on the summation of near-end crosstalk within a repeater section has not been very thoroughly investigated here but it is respon-

sible for the results described in the discussion of near-end crosstalk in a five-mile length.[11]

The relative importance of various tertiary circuits in the production of interaction crosstalk between two sections was studied for the case of near-end near-end crosstalk between two 12,000-foot lengths. It was found that the outer conductor-quads and outer conductor-sheath circuits were about equally important and that crosstalk via the quad-sheath tertiary circuit was from 20 to 30 db less. These results are about as expected since the outer conductors are the source of the

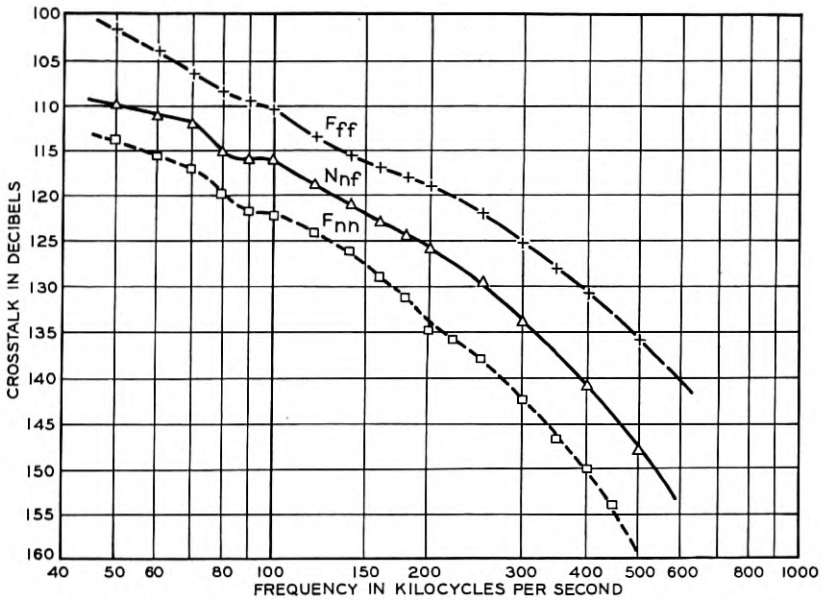

Fig. 11—Interaction crosstalk between two 12,000-foot lengths.

tertiary emf and thus the tertiary circuits involving the outer conductors should be the important ones. It is therefore permissible to consider sheath and quads as a single unit as was done in Part I.

Effectiveness of Transpositions on Far-End Crosstalk Reduction

In a long repeatered system the far-end crosstalk measured in successive individual sections inherently tends to sum up directly since all

[11] It should be noted that while Fig. 11 shows the measured values of the three types of interaction crosstalk between two 12,000-foot sections, the relative importance of the various types acting *between* repeater sections, that is, around repeaters, is not as shown there, since different correction factors have to be applied when estimating the total crosstalk at system terminals.

repeaters have practically the same phase shift and the propagation characteristics of the two coaxials are nearly identical. One way to prevent this direct addition is to transpose one section against another or one group of sections against another group along the line. In the case of unbalanced circuits these "transpositions" take the form of transformers or extra tube stages in one of the systems at repeaters, either of which will produce a 180-degree phase reversal.

If the far-end crosstalk in one transposition section is F_{l1} and that in another is F_{l2} the total in the two sections, *exclusive of interaction crosstalk between sections*, is inherently $F_{l1} + F_{l2}$. With a transposition in one coaxial at the junction the total becomes $F_{l1} - F_{l2}$. Hence, if $F_{l1} = F_{l2}$ it is possible to eliminate this crosstalk component entirely. However, due to irregularities in the cable and the practical impossibility of locating repeater points exactly, F_{l1} will not, in general, equal F_{l2} and even after transposing a small residual may remain.

This residual, however, may be negligible compared with the near-end near-end and far-end far-end interaction crosstalk components F_{nn} and F_{ff} between repeater sections (that is, around repeaters), unless transmission along the tertiary circuits from one repeater section into another is suppressed at repeater points. The interaction crosstalk tests already discussed may be used to compute this effect. However, in order to demonstrate the effectiveness of transpositions, far-end crosstalk tests were made in a 24,000-foot length with and without a transposition in one of the coaxials at the center and with various interaction crosstalk paths suppressed. The results are given in Figs. 12 to 14 and are discussed below.

To suppress entirely the interaction crosstalk between the transposed sections all tertiary circuits were shorted at the transposition point. In these measurements the tertiaries were also shorted at each end of the line in an effort to have both ends of each half of the line terminated as nearly alike as possible. The test results are given in Fig. 12.

For this condition the crosstalk measured in each half of the line is also shown. Curve AB represents the far-end crosstalk in one line section and $A'B'$ that in the other section. Curve $(AB + A'B')$ gives the results when the two sections are combined with no transposition. Curve $(AB - A'B')$ gives the results when a transformer is inserted in one coaxial at the center. (A similar set of curves are given for BA, $B'A'$, etc.)

Note that AB and $A'B'$ coincide very closely in magnitude. When combined with no transposition the crosstalk in two sections is nearly 6 db higher over the entire frequency range than in either individual

section. When combined with a transposition the crosstalk in two sections is from 13 to 27 db below either individual section over the frequency range. *Such a reduction is possible only because AB and A'B' are so nearly equal.*

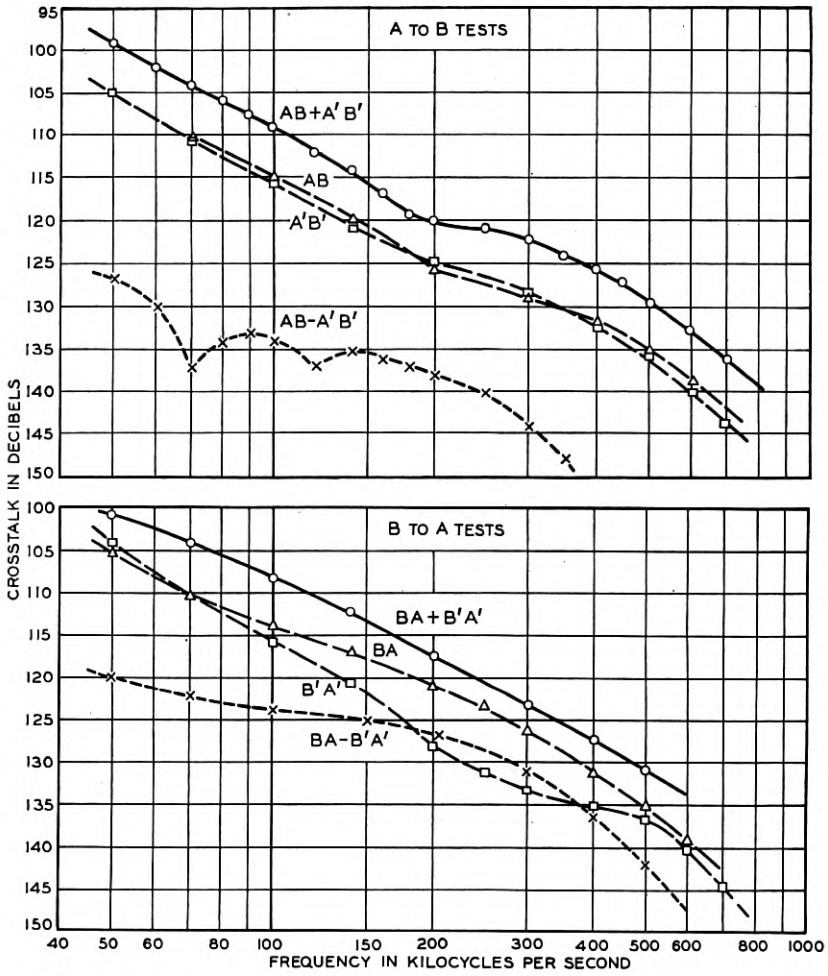

Fig. 12—Effect of a transposition on far-end crosstalk in a 24,000-foot length with all tertiary circuits suppressed at the transposition.

In contrast, BA and $B'A'$ may be seen to differ considerably from each other at the higher frequencies. As a result, the transposition is not nearly so effective in that range. The improvement at the lower frequencies where it is needed most is still about 20 db.

In order to suppress only a portion of the interaction crosstalk between two sections, measurements were made with the coaxial outer conductor-sheath circuit shorted at the transposition point thus permitting continuity of the quad-outer conductor tertiary circuit. This tertiary circuit had been shown previously to be an important

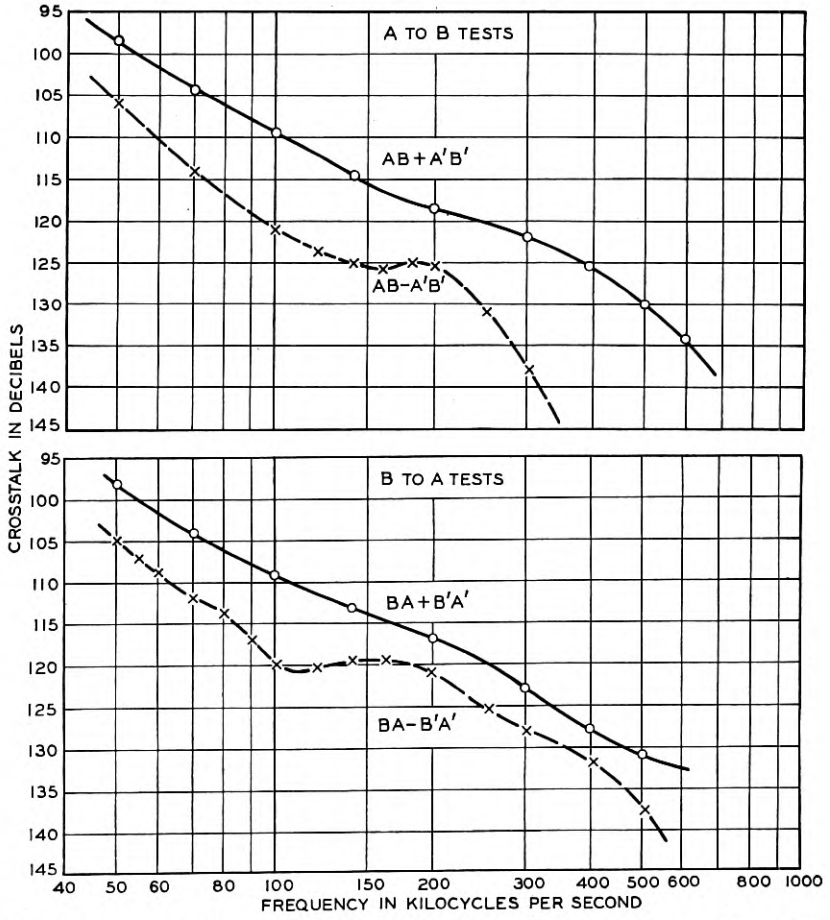

Fig. 13—Same as Fig. 12 except with quad-outer conductor tertiary circuit continuous past the transposition.

one in the production of interaction crosstalk. The measured far-end crosstalk results are given in Fig. 13.

It is at once apparent that the transposition is not so effective in this case. The crosstalk remaining after transposing is about what would be expected due to interaction crosstalk between sections via the quad-

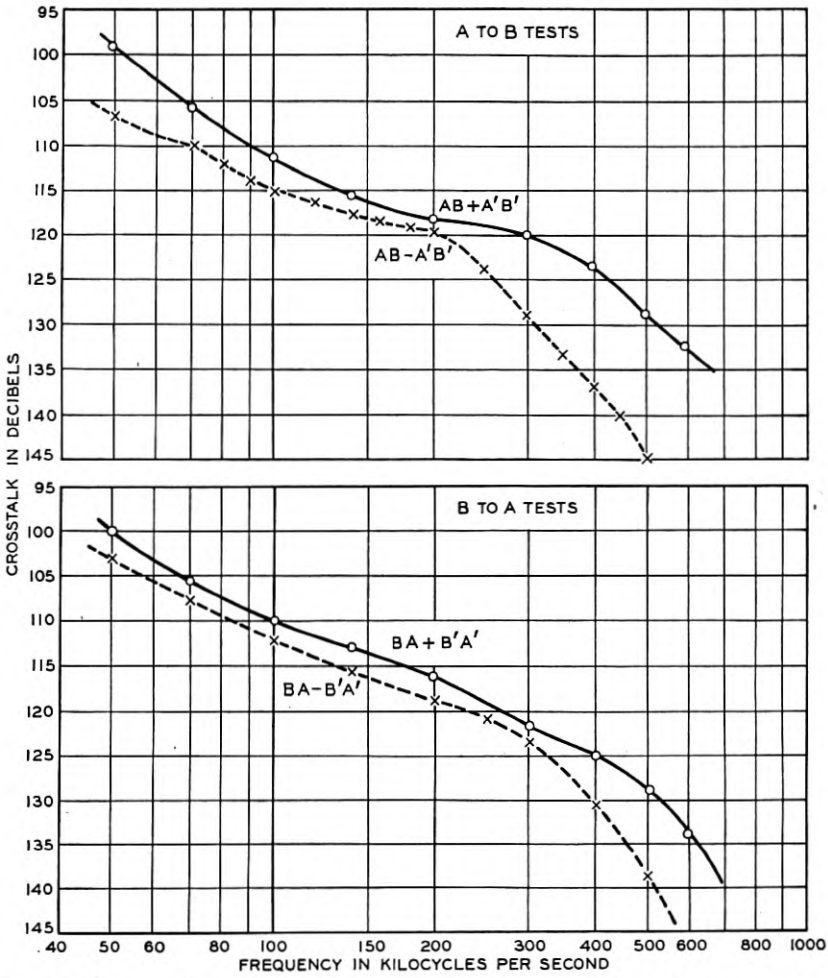

Fig. 14—Same as Fig. 12 except with all tertiaries continuous past the transposition.

outer conductor tertiary circuit.[12] However, a certain portion is also due to differences between AB and $A'B'$ (or BA and $B'A'$).

On Fig. 14 are plotted far-end crosstalk values when two 12,000-foot sections are combined with and without a transposition in one coaxial

[12] It should be noted here that these tests indicate directly the effect of a transposition *at the center* of a 24,000-foot section rather than at a *junction* between two repeater sections in a long repeatered system. If 12,000-foot repeater spacing is assumed with the transposition at the repeater point it is necessary to reduce the measured far-end far-end interaction crosstalk and increase the measured near-end near-end interaction crosstalk by an amount equal to the line loss in 12,000 feet. These corrections put interaction crosstalk between repeater sections on an output-to-output or equal level basis.

at the center and when all tertiary circuits are continuous at the transposition point and terminated at the ends. Curve ($AB + A'B'$) gives the results when the two sections are combined with no transposition. Curve ($AB - A'B'$) shows the result when a transformer is inserted in one coaxial at the junction. (A similar set of curves is given for BA, $B'A'$, etc.)

It is seen that in the 50–200 kc range there is an improvement in overall crosstalk of from 3 to 8 db due to the transposition. However, the overall crosstalk in the combined sections with a transposition is not appreciably less than that in an individual 12,000-foot section as shown by curve F_l on Fig. 8. Reference to Fig. 11 shows that this is due mainly to the far-end far-end interaction crosstalk between the two sections which is unaffected by the transposition.

The results shown in Fig. 12 give some indication of the extent to which far-end crosstalk may be reduced by means of a transposition, *provided interaction crosstalk between sections is entirely suppressed.* As illustrated in Figs. 13 and 14 a transposition at a repeater point is not nearly so effective if the interaction crosstalk is not suppressed.

ACKNOWLEDGMENT

The authors are greatly indebted to Mr. John Stalker and the staff at the Princeton, New Jersey, repeater station of the American Telephone and Telegraph Company and to Mr. William Bresley and Mr. Norman Mathew of the New Jersey Bell Telephone Company, for their cooperation and assistance in the Princeton tests.

Compressed Powdered Molybdenum Permalloy for High Quality Inductance Coils *

By V. E. LEGG and F. J. GIVEN

Molybdenum-Permalloy is now produced in the form of compressed powdered cores for inductance coils. Its high permeability and low losses make possible improved coil quality, or decreased size without sacrificing coil performance. Its low hysteresis loss reduces modulation enough to permit application where large air core coils would otherwise be required.

Introduction

THE introduction of loading coils in the telephone system at about the turn of the century brought special demands on magnetic and electrical properties of core materials, and set in motion investigations which have had wide influence on the theoretical and practical aspects of ferromagnetism. The first step in this development led to cores of iron wire, which sufficed for loading coils on circuits of moderate length.[1] With the development of telephone repeaters and the extension of circuits to transcontinental length some twenty-five years ago, there arose need not only for loading coils, but also for network coils, which would have high stability with time, temperature and accidental magnetization. Magnetic stability was at first secured [2] by employing iron wire cores provided with several air gaps. Later, commercial and technical considerations led to a core structure made from compressed insulated powdered material, first electrolytic iron [3] and later permalloy powder.[4] This type of core is mechanically stable; it introduces in an evenly distributed fashion the requisite air-gaps, while avoiding undesirable leakage fields; and it sub-divides the magnetic material so as to reduce eddy-current losses. Although other means have been suggested,[5, 6] no way has yet been devised which provides these features so well and at so low a cost as the compressed powdered type of core.

Loading coil cores made from electrolytic iron powder generally satisfied the stability requirements for long lines, but on account of their low magnetic permeability they were large and costly. The search for materials with higher permeability and lower hysteresis loss

* Presented at Winter Convention of A.I.E.E., New York, N. Y., January 22–26, 1940.

led to permalloy [7] which, by 1925, had been produced in powdered form and fabricated into cores. This development provided coils for voice frequency applications (loading coils and filter coils) which were cheaper and yet superior electrically to those made from electrolytic iron. These coils became available at a time when the telephone plant was undergoing a very large extension of loaded cables. As a result, large economies in cost and space were realized in the more than six million coils involved in this plant expansion.

Iron powder, and later permalloy powder, ground to a finer size and diluted to lower permeability than used in loading coils, also found application in coils for oscillators, filters and networks of multiplex carrier telephone and telegraph systems employing frequencies up to 30 kc.[8] and in receivers for transoceanic radio telephone communication employing frequencies up to approximately 60 kc.[9] Permalloy powder improved the electrical characteristics—particularly modulation—of coils for use in high frequency circuits, because of its low hysteresis losses.

Continued research for a powdered material having still better intrinsic properties has recently made available new compressed powder cores which permit further important gains in coils for voice-frequency circuits and in coils for high-frequency carrier system applications. The latter take on considerable significance at this time because they play an important part in making practical for commercial use the new broad-band carrier telephone systems intended for use on existing open-wire and cable lines and on new types of cable. Again, therefore, the advent of a new core material is well-timed to be of assistance in further growth of the telephone system.

The development of this core material was based on the discovery [10] that the addition of a small percentage of molybdenum to permalloy increases its permeability and electrical resistivity, and decreases its eddy current and hysteresis losses. Decreased losses are necessary for improvements and economies for both voice and carrier frequency operation. The increased permeability of this alloy is essential for the improvement of voice-frequency coils. It is readily reduced to the proper values for high-frequency coils by diluting the powdered magnetic material with insulating material before compressing into core form. In this development many problems of alloy embrittlement, pulverization, insulation and heat treatment had to be solved both on a laboratory and factory scale. The alloy composition finally selected as giving the best combination of desirable properties, contains approximately 2 per cent molybdenum, 81 per cent nickel, and 17 per cent iron, and is designated as 2–81 molybdenum-permalloy. This new

alloy is manufactured commercially by the Western Electric Company for use in loading coils and filter coils.

Physical and Magnetic Characteristics

The raw materials and necessary embrittling agents [11] are melted together and cast into ingots which are rolled to develop the desired grain structure. The density of this alloy is 8.65 gm/cm³. The brittle material is pulverized to the desired fineness and finally annealed to soften the alloy particles before insulation and pressing into core form.

The distribution by weight of the particle sizes of a sample of 120-mesh powder is given in Fig. 1, showing a root mean square size of 50

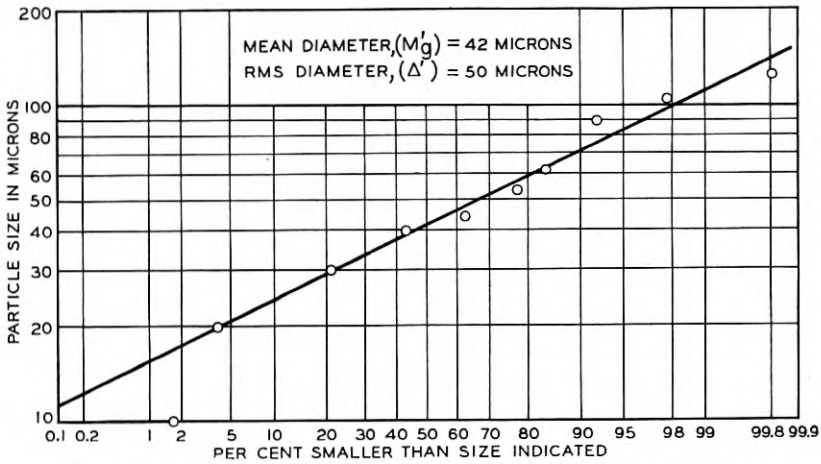

Fig. 1—Distribution of particle size of 120-mesh powder, by weight.

microns.[12] Since the effective resistance of a coil due to eddy-current losses in its core is proportional to the mean square particle diameter,[13] it can be decreased when desired by the use of more finely pulverized material.

The problem of insulating 2–81 molybdenum-permalloy powder is to coat the particles with a minimum thickness of a material which will not break away during the pressing operation, which will not fuse and flux the magnetic particles together during the core heat treatment, which will prevent the flow of eddy currents between metallic particles, and which will be chemically inert throughout the lifetime of the magnetic core. The difficulty of the problem will be appreciated from the fact that the separation between adjacent particles of a core of 125 permeability is approximately equal to the wave-length of visible

light (0.5 micron). This thickness of insulating film may be shown to be approximately $\frac{rt}{300p}$, where r is the percentage of insulating material by volume, p is the packing factor of the magnetic material, and t is the r.m.s. particle diameter (assuming spherical particles).

A new type of ceramic insulation has been introduced with these cores which fulfills the above requirements and which is more inert than the previous type. This new insulating material is free from water soluble residue. It thus eliminates the final washing treatment which was required with the earlier type.

For applications where a low permeability is desired, non-magnetic powder is added to further dilute the magnetic material. The permeability of the finished core depends largely on the quantity, particle size, and thoroughness of admixture of non-magnetic powder. Various attempts to derive theoretical relations between core permeability and dilution have been made,[14,15] but they generally fail in some detail. An empirical representation of this relationship is found to be

$$\mu = \mu_i^p \qquad \text{or} \qquad \log \mu = p \log \mu_i,$$

where μ_i is the intrinsic permeability of the magnetic material, and p is the packing factor, or fraction of the core volume occupied by magnetic material. This equation is found to be valid for a wide range of dilution, effected either by adding insulating material or by reducing the load during core compression. However, the intrinsic permeability must be determined experimentally for each type of particle, size distribution, method of admixture of non-magnetic powder, and annealing process. Figure 2 shows curves of permeability and percentage diluting material vs. metallic packing factor. A permeability of 125 has been selected for most loading coil cores, while permeabilities of 14 and 26 have been chosen for two important types of high-frequency filter coils.

A pressure of 100 tons/sq. in. is employed in forming molybdenum permalloy cores, to attain proper density and mechanical strength. The effect of pressure on core density and strength is shown in Fig. 3 for cores having 2.5 per cent dilution. The tensile strength of diluted cores is decreased somewhat; for example, cores with 25 per cent non-magnetic materials have a tensile strength of about 250 lbs./sq. in.

In annealing the compressed cores to remove stresses incident to pressing, it has been found that the insulating material remains intact at a considerably higher temperature if oxygen is excluded. An improved annealing treatment has therefore been introduced by which

the cores are heated in an atmosphere of hydrogen, with the attainment of high core permeability and low hysteresis loss. The ability of the insulating material to withstand such a heat treatment testifies

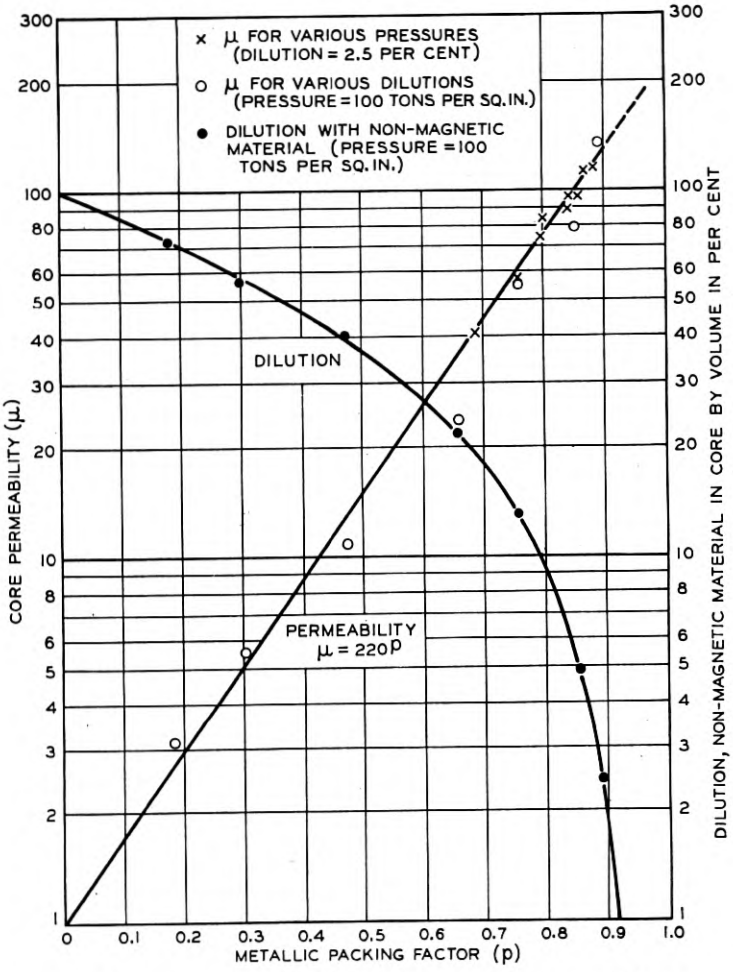

Fig. 2—Relation between metallic packing factor, permeability and percentage dilution.

to its extreme stability and recommends it in preference to organic materials.

An essential core requirement of precision inductance coils is that the permeability remain unaltered during the life of the coil. The greatest difficulty with cores having no air gaps is the large and more or

less permanent shift of permeability due to accidental strong magnetization. Such variations have recently been overcome to a degree in continuous cores made of hard rolled nickel-iron alloy sheet,[6] but they have been found to be rather large immediately after strong magnetization, decreasing slowly to tolerable limits only after two or three days. With compressed powdered molybdenum-permalloy cores, permeability shift due to strong magnetization is remarkably small even within a fraction of a minute after the magnetization is released, and any

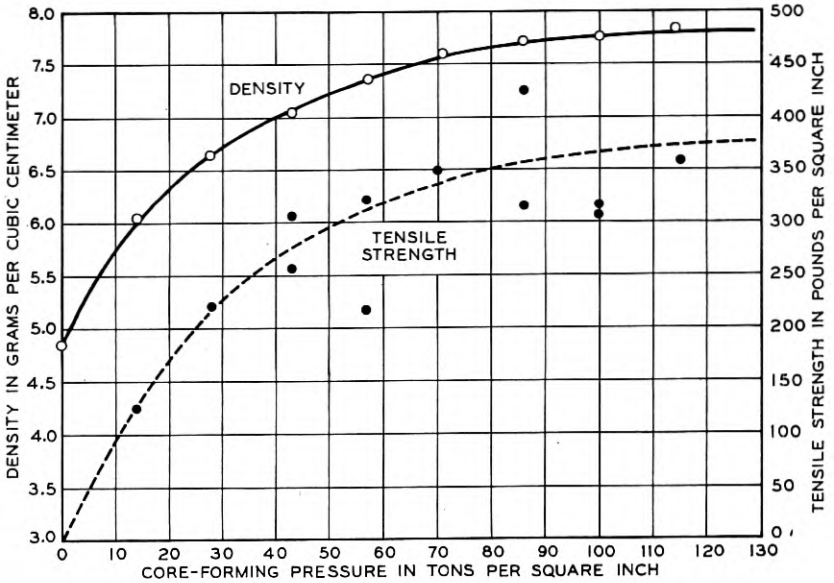

Fig. 3—Effect of core forming pressure on density and tensile strength.

further drift of permeability with time is negligible. In typical cores of the new material, the shift in permeability after strong magnetization is less than 0.2 per cent for cores of permeability 125, and less than 0.05 per cent for cores of permeability 14. Figure 4 shows the residual effect of the application and removal of various magnetizing forces on cores of both these permeabilities.

When a direct current is superposed on an alternating current in the windings of a coil, the inductance is altered because the magnetic field set up by the direct current modifies the core permeability. Figure 5 shows the effect of superposed d-c. fields on the permeability of 2–81 molybdenum-permalloy powder cores of various permeabilities.

A further important core property is the constancy of permeability with respect to flux density B. This is of particular importance in

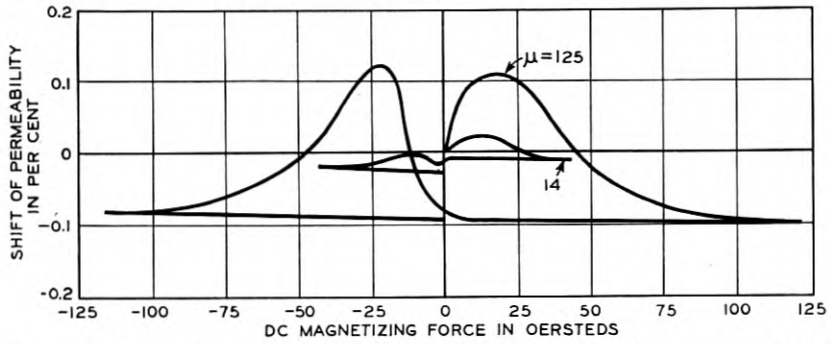

Fig. 4—Residual effect of d-c. magnetization on initial permeability—measured three minutes after release of direct current.

precision filters, to insure that changes in transmission level do not produce serious alterations in the frequency discrimination characteristics. Figure 6 shows the superiority of the new material over the earlier permalloy.

A new requirement for cores has been introduced by quartz crystal filters used in wide-band carrier systems. In order to secure the necessary precision in this type of filter, measures have to be taken to prevent departures from the initial frequency adjustment due to changes in core permeability ordinarily occurring with room temperature changes. Extremely small temperature coefficients of permeability have now been achieved by adding to the new 2–81 molybdenum-permalloy powder a very small percentage of special permalloy powder having a molybdenum content of about 12 per cent. Such an alloy has a non-magnetic or Curie point close to room temperature, and for a

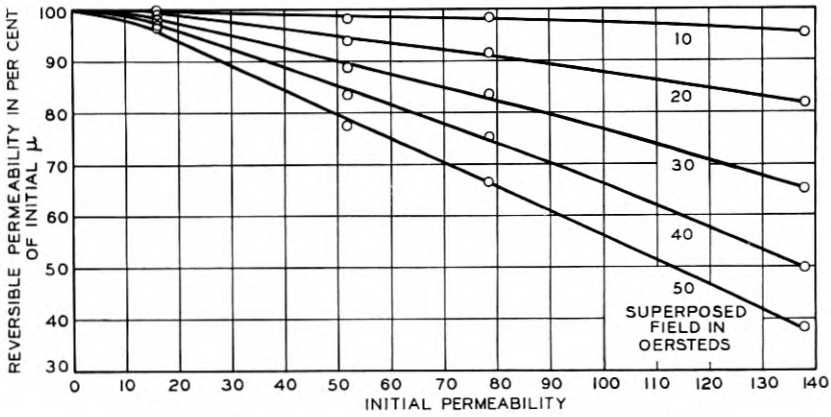

Fig. 5—Effect of superposed magnetization on permeability.

small temperature range just below its Curie point, it has a negative temperature coefficient several hundred times as large as the positive coefficient of 2–81 molybdenum-permalloy. By choosing suitable compositions and percentages of such compensating alloys, the net temperature coefficient of permeability of a core can be adjusted to any reasonable value, positive or negative, over a desired temperature range. Figure 7 shows a permeability vs. temperature curve for a

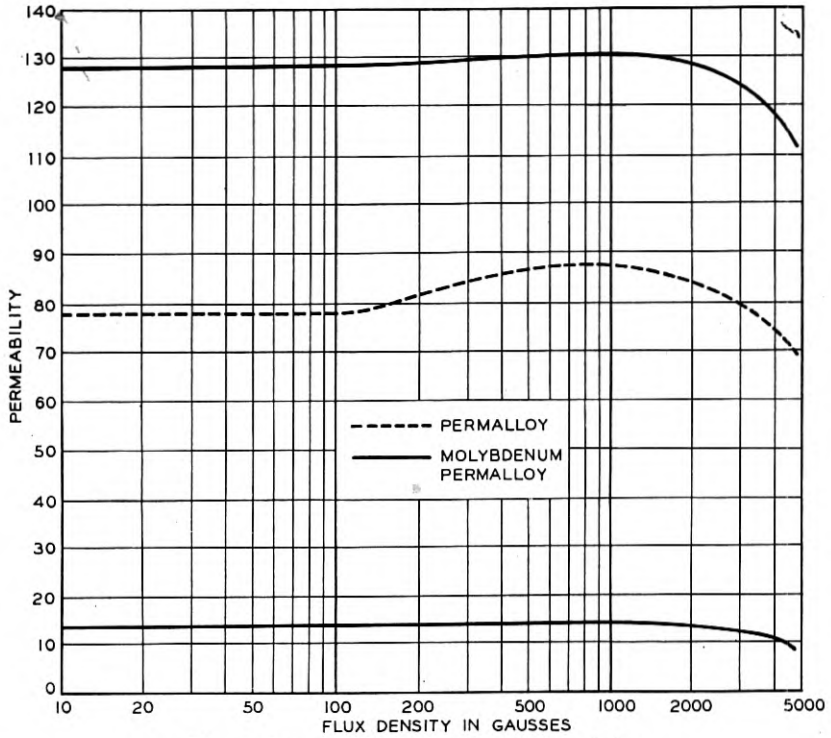

Fig. 6—Permeability-induction characteristics.

core stabilized to give a small negative coefficient, compared to a similar curve for a core not stabilized.

Core Losses

The desirability of core materials increases in general as their loss characteristics decrease. Low total eddy-current and hysteresis losses give low contributions to attenuation. Hysteresis loss is frequently of especial importance because it appears fundamentally as a resistance which varies with coil current, and because it incidentally generates harmonic voltages. A low value of hysteresis loss thus

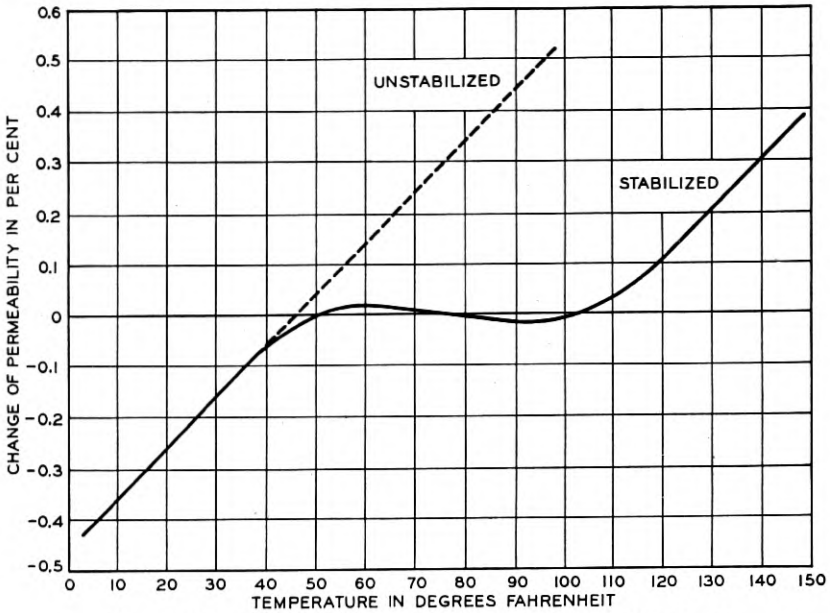

Fig. 7—Effect of temperature on permeability of stabilized and unstabilized cores.

simplifies circuit problems arising from resistances which vary with energy level, and it avoids troublesome modulation conditions.

The total resistance per unit inductance arising from eddy-current and hysteresis losses may be expressed as [16]

$$R_m/L = \mu(aB_m + c)f + \mu ef^2,$$

where the symbols are as given in the Appendix.

Table I gives the loss coefficients for various core materials, and for

TABLE I

Loss Coefficients of Powdered Core Materials

Material	μ	Hysteresis		Residual		Eddy Current	
		$a\times10^6$	$\mu a\times10^3$	$c\times10^6$	$\mu c\times10^3$	$e\times10^9$	$\mu e\times10^6$
Grade B Iron	35	49	1.7	109	3.8	88	3.1
Grade C Iron	26	81	2.1	139	3.6	31	0.8
81 Permalloy	75	5.5	0.41	37	2.8	51	3.8
	26	11.5	0.30	108	2.8	27	0.7
2–81 Molybdenum Permalloy	125	1.6	0.20	30	3.8	19	2.4
	26	6.9	0.18	96	2.5	7.7	0.2
	14	11.4	0.16	143	2.0	7.1	0.1

2–81 molybdenum-permalloy insulated to several permeabilities. The low loss coefficients of the new material as compared with the best previous materials are of importance from two standpoints. First, core permeabilities as much as 50 per cent greater can be now utilized in coils without increasing the total core loss resistance. Second, by utilizing the same permeabilities, core loss resistances about 60 per cent smaller can be obtained.

In many coil design problems, harmonic generation or modulation assume controlling importance. The modulation factor m which denotes the ratio of the generated third harmonic to the applied voltage may be expressed as follows [17]

$$m = E_3/E_1 = 3\mu a B_m/10\pi.$$

The low values of a obtained with the new material yield values of m that are about 6 db and 20 db lower than possible with powdered permalloy and electrolytic iron cores, respectively.

The wide range of core permeability available with this new material permits a ready choice of the proper values of permeability and core size to suit any particular needs. In the usual design problem the following main requirements must be considered, in addition to providing the desired inductance.

1. D-C. Resistance, R_c.
2. Coil quality factor, $Q = \omega L/(R_c + R_m)$.
3. Modulation Factor, m.
4. Coil size (which depends directly on core size).

These requirements can not be satisfied independently however, as fixing any two of them automatically fixes the values of the others. In each case, a particular value of core permeability is required for the proper fulfillment of the conditions. Appendix I lists the formulae essential to the determination of these factors. Although these formulae imply an entire freedom of choice of core permeability and size, it becomes necessary for practical purposes to standardize on a limited number of values of permeability and a limited number of sizes of core. It is possible by the proper choice from these types to approach rather closely to an ideal solution for each problem.

Improved Designs of Loading Coils

A study of alternate ways of utilizing the advantages offered by the new material in coils for voice frequency loaded cables showed that the greatest immediate benefit to the telephone plant would accrue from making the new coils substantially duplicate performance characteristics of previous designs. The new designs chosen are in fact better

in most respects and have been made approximately 50 per cent smaller in volume by using a molybdenum-permalloy core with a nominal permeability of 125. Table II summarizes data comparing

TABLE II

COMPARATIVE SIZE AND WEIGHT DATA OF TYPICAL NEW AND SUPERSEDED COILS

Type of Coil	Type of Compressed Powdered Core	Inductance (Henrys)	Coil Volume (Cu. In.)	Coil Weight (Lbs.)
Small Exchange Area	Permalloy	0.088	2.5	0.4
"	Molybdenum Permalloy	0.088	1.5	0.2
Program Circuit	Permalloy	0.022	11.8	1.6
"	Molybdenum Permalloy	0.022	4.4	0.6
Toll-Side Circuit	Permalloy	0.088	13.5	1.7
"	Molybdenum Permalloy	0.088	5.1	0.7

the electrical characteristics, sizes, and weights of coils commonly used on exchange and toll cables. Figure 8 shows the improved

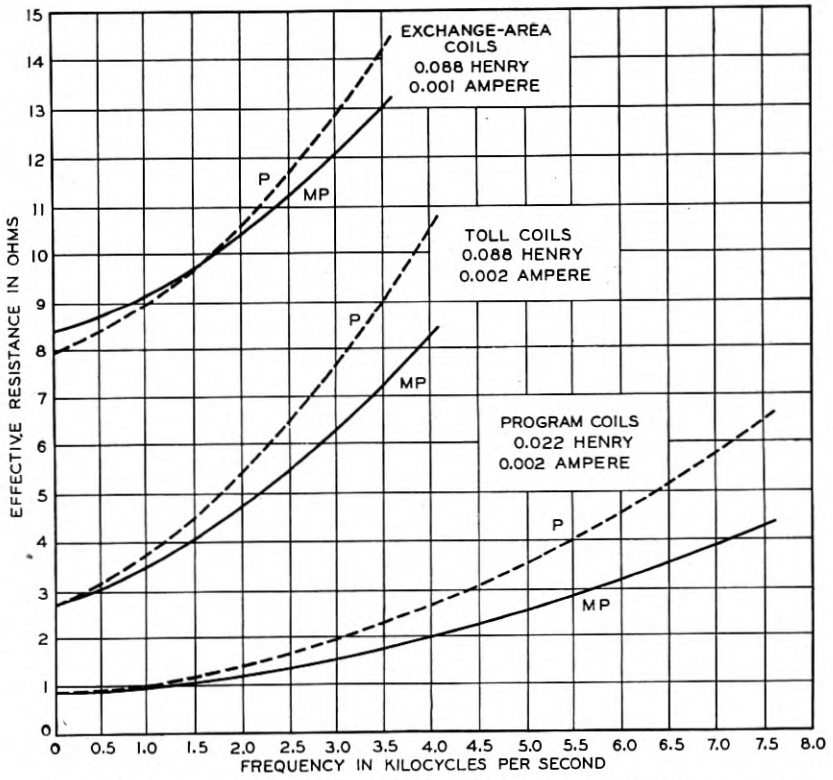

Fig. 8—Effective resistance-frequency characteristics of typical loading coils.

resistance-frequency characteristics for typical coils. Figure 9 shows the improved telegraph flutter [18] characteristics of a commonly used toll type coil.

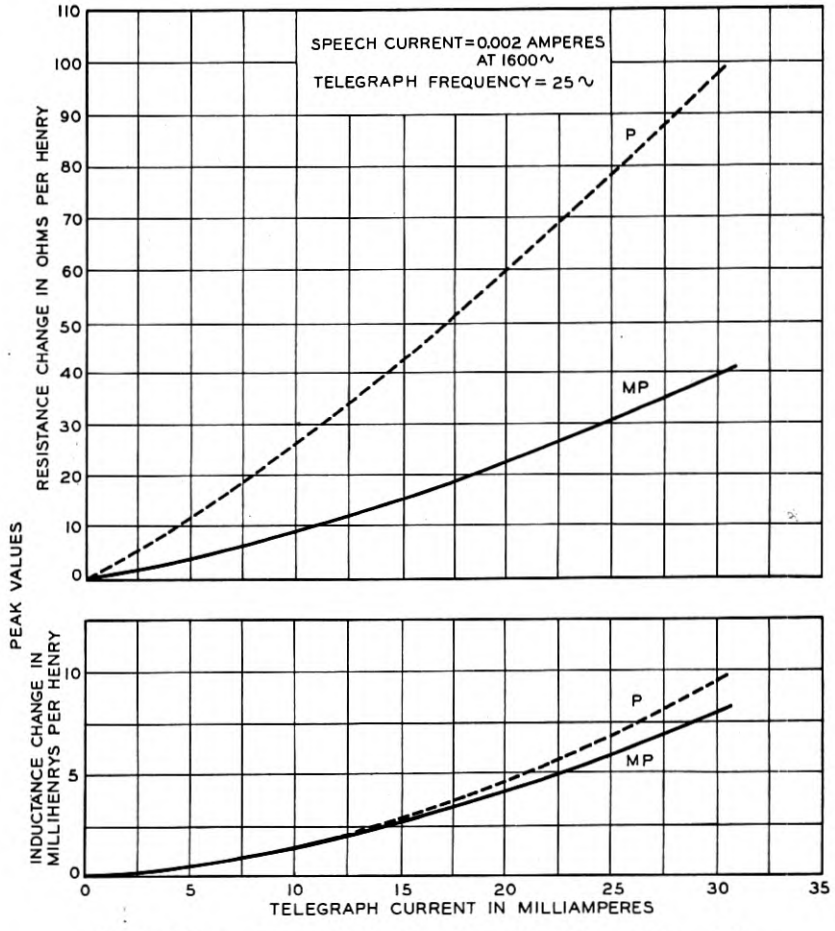

Fig. 9—Flutter characteristics of typical toll loading coil; P—with permalloy core, MP—with molybdenum-permalloy core.

Figure 10 pictures the reduction in size of cores and coils which are commonly used in toll and exchange area circuits. In the preparation for commercial manufacture of the smallest of the new coils, a difficult problem in the development of winding machinery was involved because of the small dimensions of the hole in the finished coil. This problem has been successfully solved by the Western Electric Company.

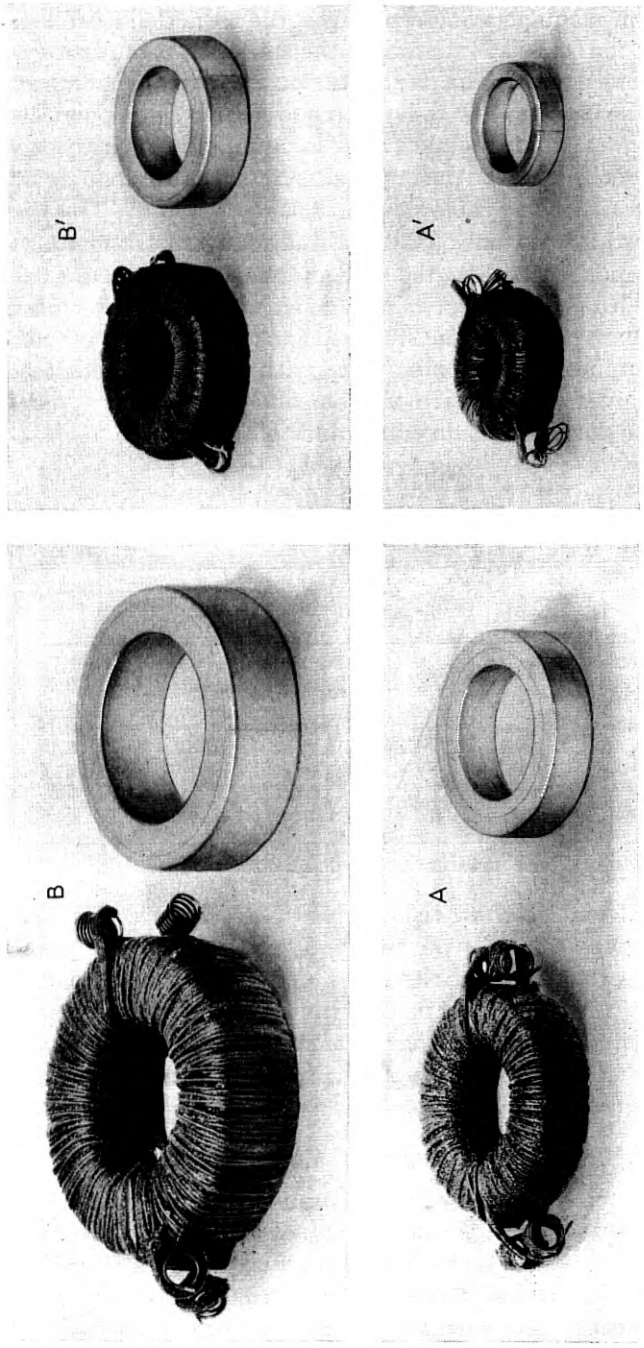

Fig. 10—Comparative sizes of the new molybdenum-permalloy (A, A') and the superseded permalloy (B, B') cores and coils: left—program circuit loading coils; right—exchange area loading coils.

Aside from manufacturing economies, the reduction in coil size is of importance to the Bell System from the standpoints of plant construction and installation. The size reduction is particularly important in instances where only a few coils are required at a given point, as is the case in program circuit loading. It is now practicable to enclose as many as six coils, even of the larger size employed for loading program circuits, directly in the cable splice at a loading point. This dispenses with the need for conventional cases, and reduces both manufacturing and installation costs. In the field of small complements using the inexpensive lead sheath type of case construction, it is now possible to furnish as many as 100 of the small exchange area coils whereas 15 was the maximum number accommodated with the superseded coil design. Table III gives comparative weights and volumes of typical cases provided for potting exchange area and toll type coils.

TABLE III

COMPARATIVE DATA ON REPRESENTATIVE CASES FOR NEW AND SUPERSEDED LOADING COILS

Type Cable	Size of Complement	Type Coil or Unit	Approx. Volume Cu. Ft.	Approx. Weight Lb.
Exchange Area	200	Permalloy	1.7	480
"	"	Molybdenum-Permalloy	0.8	350
Program-Toll	6	Permalloy	0.15	70
"	"	Molybdenum-Permalloy	0.11	50
Toll	50 units*	Permalloy	5.5	1350
"	"	Molybdenum-Permalloy	3.5	950

* A unit consists of one phantom and two side circuit coils.

Figures 11 and 12 show the comparative sizes of typical steel and lead sleeve type cases for the new and superseded coils. In Fig. 13, the midget proportions of the latest design of case for plotting 100 exchange area coils are contrasted with those of the case utilized up to 1922 containing only 98 exchange area coils. The reduction in size of cases for this type of coil is of particular importance in larger cities where underground vault space is at a premium.

IMPROVED INDUCTANCE COILS FOR FILTERS

The trend of development of toll transmission circuits is now very definitely toward multiple channel carrier systems utilizing a much wider frequency band than has heretofore been employed on wire circuits.[19,20] These systems involve extensive use of selective or equalizing networks at terminals and at repeater stations in order to

obtain proper separation of the frequency bands of the various channels or to insure suitable transmission properties of the individual channels. While these networks involve coils, condensers and crystals, it is frequently the case that their size, cost and performance are determined chiefly by the quality factor Q of the inductance coils. This follows from the fact that Q values of coils are usually considerably

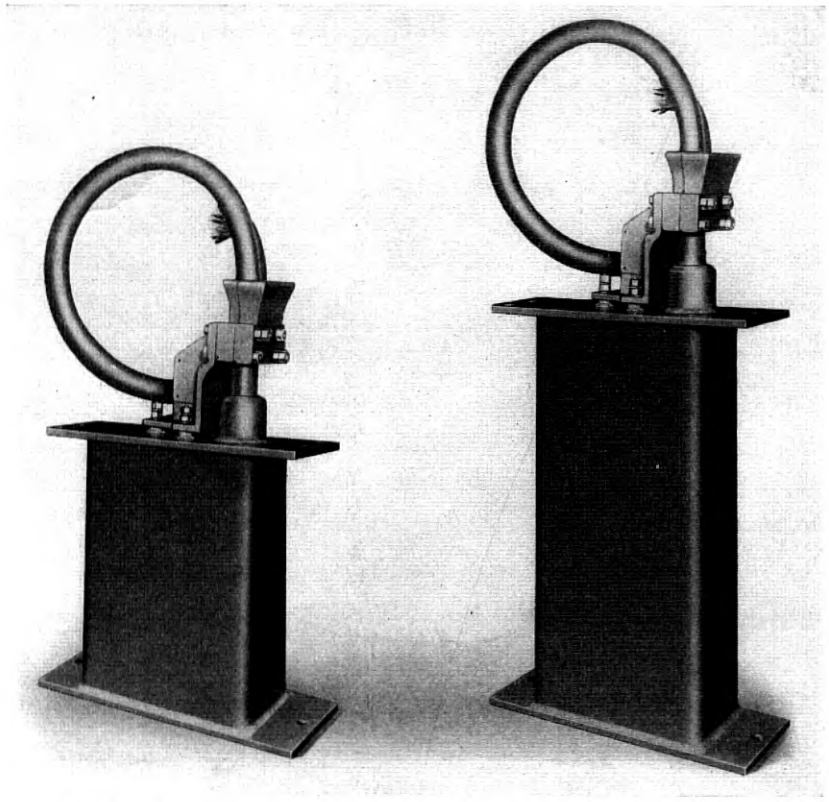

Fig. 11—New and superseded cases containing 200 exchange area coils.

lower than those obtainable readily in condensers and crystals. Accordingly, it is very desirable to have as high a value of Q as possible economically. In addition, such coils must have low hysteresis resistance to limit modulation, and a low temperature coefficient of inductance to secure stability of attenuation or impedance characteristics of the filters and networks.

Due to the improvements in these respects, molybdenum-permalloy core coils can be used quite extensively in new types of carrier tele-

phone systems. In such systems for existing lines and cables, as well as for projected new types of cables, those filters are of key importance which separate individual message channels in the frequency range from 3 to 108 kc. By using coils of powdered molybdenum-permalloy insulated to permeabilities of 14 or 26, valuable economies in space and cost of filters are realized.[21,22] Figure 14 shows a typical coil employing a 14 permeability core designed for use in one of these channel filters having its transmitted band in the vicinity of 108 kc,

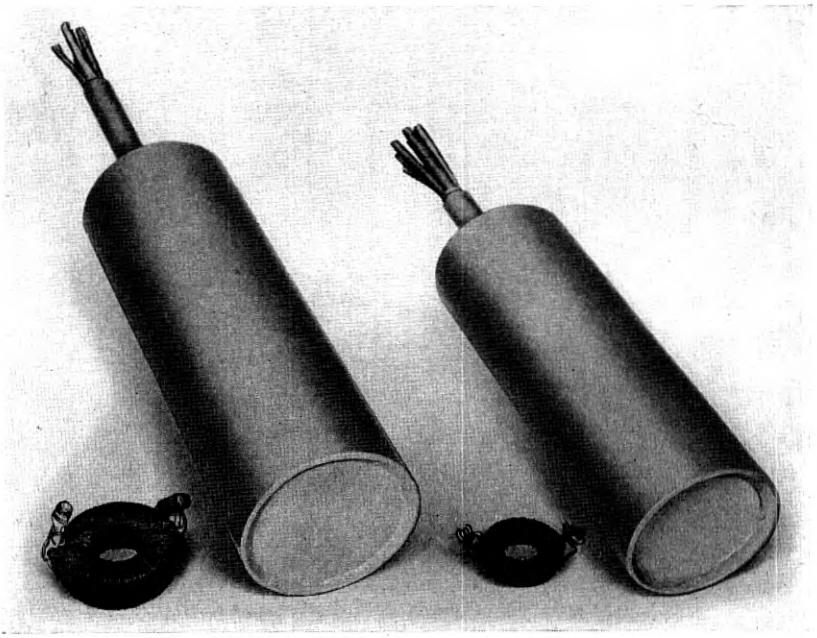

Fig. 12—New and superseded cases containing six program loading coils.

together with a shielded solenoidal air core coil which might be employed for the same purpose. The molybdenum-permalloy coil has a Q at 100 kc about twice that of the air core coil, yet it occupies approximately 1/10 as much space. The third order modulation products are approximately 80 db below the level of the normal channel currents. This is considered to be tolerable from the standpoint of interchannel crosstalk on circuits used for one-way transmission. An inductance-temperature coefficient of about -20×10^{-6} per degree F. has been chosen to compensate for the positive capacity-temperature coefficient of associated condensers.

In Fig. 15 data are presented illustrating the Q-frequency characteristics that can be obtained on typical coil designs using the new material in the frequency range from 300 cycles to 200 kc. The characteristics shown apply to coils wound on three sizes of cores that

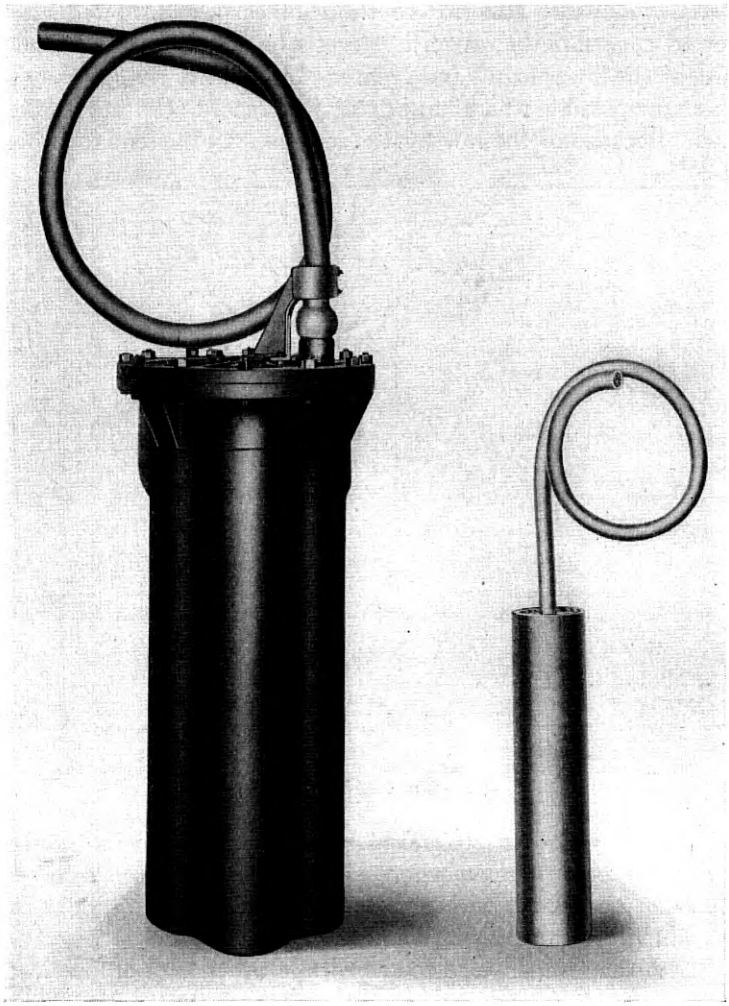

Fig. 13—Comparative size of equivalent 1939 and 1922 cases.

are suitable for use in this range. For comparison, similar characteristics are also included for coils using cores of equal size but made of permalloy and electrolytic iron powder. These data include all effects on Q resulting from winding capacities and losses, which have

been made tolerably small by suitable choice of insulating materials, stranding of conductor and configuration of winding.

Conclusion

Compressed powdered cores of 2–81 molybdenum-permalloy have properties which are superior to those of earlier powdered cores in respect to permeability range, hysteresis loss and eddy current loss. Because of the lower losses and greater permeability range, inductance coils are now possible which have greatly increased Q values for a given volume. Because of the low hysteresis losses and attendant lowering

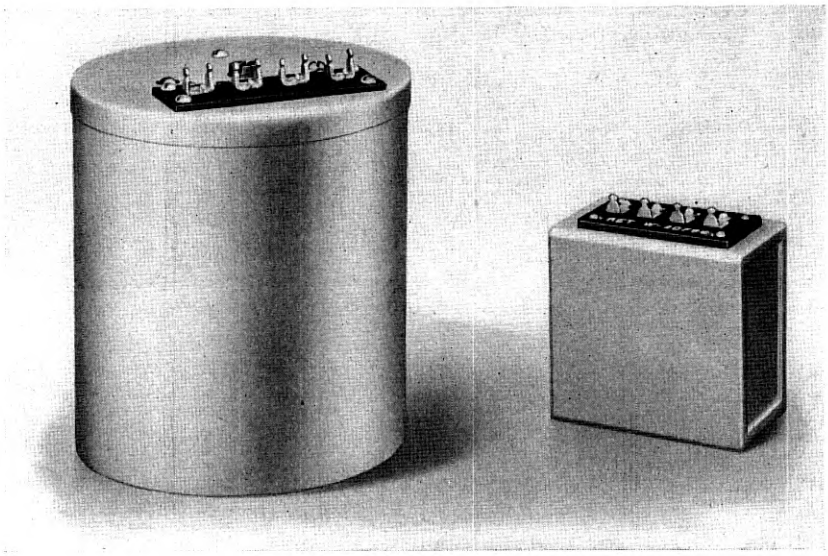

Fig. 14—Comparative size of non-magnetic core and molybdenum permalloy core filter coils.

of modulation effects to tolerable levels, magnetic core inductance coils can now be employed where non-magnetic core coils have previously been necessary, with a very great increase in Q values for a given volume. Temperature coefficients can now be obtained which are equal to the temperature coefficients of other high grade electrical elements such as mica condensers and quartz crystals. Moreover, the ability to make the temperature coefficient of coils negative or positive at will permits the attainment of remarkable stability in resonant combinations of coils and condensers. Two important applications have been made in the field of communication apparatus. For voice-frequency circuits, new loading coils of improved quality and reduced

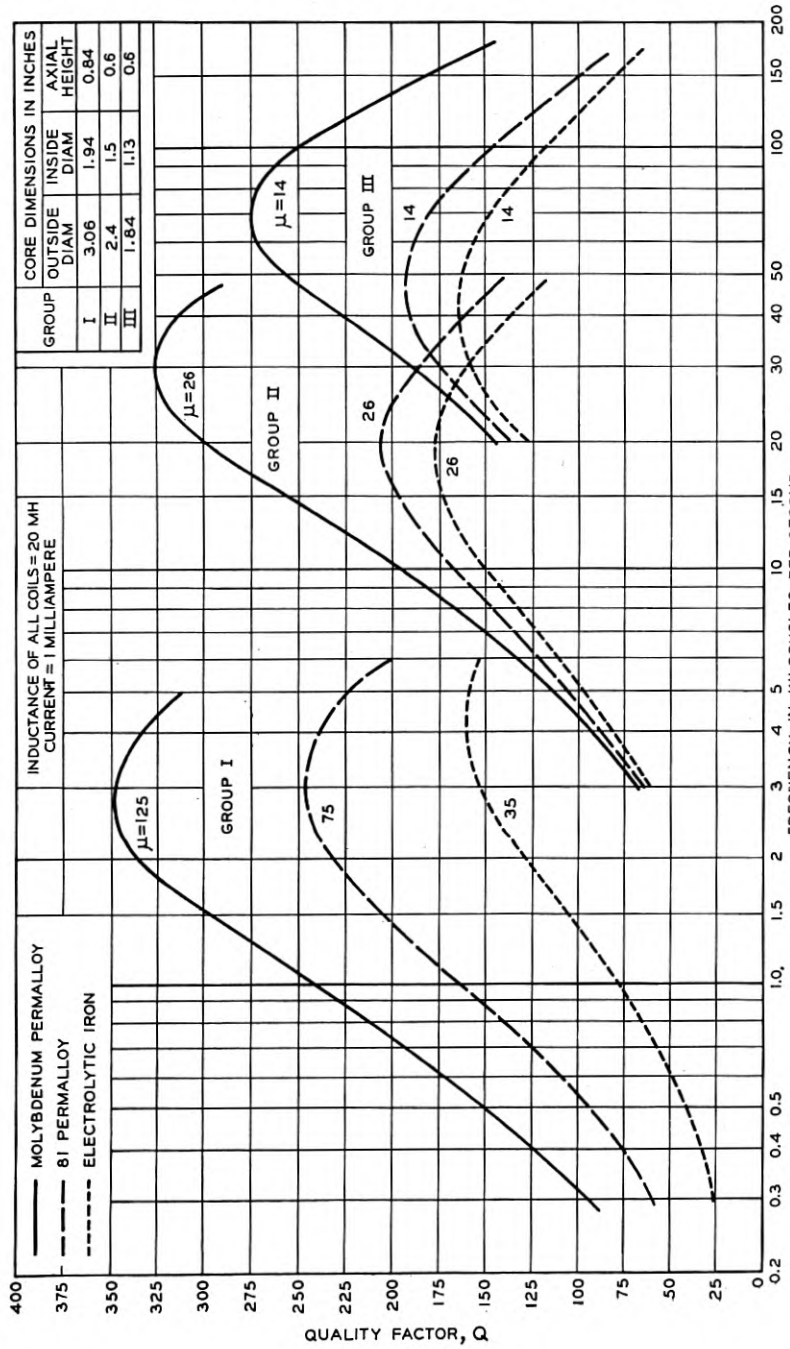

Fig. 15—Comparative Q-frequency characteristics on typical filter coils using new and superseded materials.

size have been standardized. For broad band carrier systems, the compactness of channel separating filters is largely due to the volume economies introduced by molybdenum-permalloy coils.

APPENDIX

The following formulae illustrate the essential steps in selecting the size and permeability of an annular core best suited for a coil having a desired set of characteristics. For sake of simplicity, they assume throughout a coil of arbitrarily chosen proportions, in which all dimensions bear fixed ratios to the mean core diameter, and approximating those that are practicable from a manufacturing standpoint. The core is assumed to be rectangular in section. It is assumed that wire of any diameter can be used and that the winding efficiency is independent of the wire diameter. Any inductance due to air space outside of the core is neglected. Because of these simplifications, the expressions are of somewhat restricted applicability. However, they yield solutions for optimum permeability and corresponding values of Q and core size which are sufficiently accurate for most practical purposes.

The inductance in henrys due to a core of permeability μ, and mean diameter d cm., wound with N turns of wire is

$$L = \frac{3}{8} N^2 \mu d \times 10^{-9}. \tag{1}$$

If the coil is wound with wire of resistivity ρ_c ohm-cm., with winding efficiency s (i.e., the ratio of copper area to total available winding area), the direct current resistance in ohms will be

$$R_c = \frac{19 \rho_c L \times 10^9}{s \mu d^2}. \tag{2}$$

The maximum flux density due to sine wave measuring current of effective value I amperes is

$$B_m = \frac{8I}{5} \sqrt{\frac{\mu L \times 10^9}{3d^3}}. \tag{3}$$

In terms of the hysteresis loss coefficient $k_2 = \mu a$, the modulation factor thus becomes [17]

$$m = \frac{E_3}{E_1} = \frac{3 k_2 B_m}{10\pi} = \frac{4 k_2 I}{25\pi} \sqrt{\frac{3\mu L \times 10^9}{d^3}} \tag{4}$$

or

$$d^3 = 3\mu L \times 10^9 \left(\frac{4 k_2 I}{25 \pi m} \right)^2. \tag{5}$$

When the core permeability μ is reduced by dilution of a given material, the hysteresis and residual loss coefficients a and c vary so as to make the products $\mu c = k_1$ and $\mu a = k_2$ approximately constant, as may be seen by reference to Table I. The core loss resistance in ohms at frequency f cycles per second may therefore be expressed with reasonable accuracy as

$$(6) \qquad R_m = Lf(k_1 + k_2 B_m + \mu ef) = Lf\left(k_1 + \frac{10m\pi}{3} + \mu ef\right),$$

where the eddy current coefficient e depends upon the particle diameter t, and the alloy resistivity ρ being proportional [13] to t^2/ρ.

The coil quality factor is thus

$$(7) \qquad Q = \frac{2\pi fL}{R_c + R_m} = \frac{2\pi f}{\frac{19\rho_c \times 10^9}{s\mu d^2} + f\left(k_1 + \frac{10m\pi}{3} + \mu ef\right)}.$$

Case I: If the value of m is fixed and d and R_c can be freely chosen, it is desirable to know the value of μ which will yield the highest possible value of Q. By substituting in (7) the value of d^2 obtained from (5) and setting the derivative with respect to μ equal to zero, the following is obtained for the optimum permeability:

$$(8) \qquad (\mu')^8 = \frac{52.5 \times 10^{16}\rho_c^3 m^4}{e^3 s^3 f^6 k_2^4 I^4 L^2}.$$

The corresponding values of d and R_c can be obtained from equations (5) and (2). The corresponding value of Q, which is the greatest obtainable under these conditions is

$$(9) \qquad Q'_{\max} = \frac{\pi}{\frac{5m\pi}{3} + \frac{k_1}{2} + \frac{4}{5}\mu' ef}.$$

If a smaller value of Q than that obtained from (9) is acceptable, equations (7) and (5) can be solved simultaneously for d and μ. A smaller value of μ than that obtained from (8) and a correspondingly smaller value of d will result.

Case II: If modulation is unimportant and the hysteresis loss resistance is negligible in comparison with other component losses, then d and μ can be selected without regard to modulation. Equation (7) can be differentiated directly and solved for the permeability required to yield the maximum value of Q. This optimum permeability is

$$(10) \qquad (\mu'')^2 = \frac{19\rho_c \times 10^9}{sed^2 f^2}.$$

The corresponding value of Q is

$$Q''_{\max} = \frac{2\pi}{k_1 + 2\mu''ef}. \tag{11}$$

The values of $Q''_{\max}$ and μ'' depend on the value of d chosen. For a desired value of Q, (11) can be solved for μ'', and (10) for the corresponding diameter.

In any case, the ideal wire size has a cross-sectional area of conductor equal to

$$0.15sd^2\sqrt{\frac{\mu d \times 10^{-9}}{L}} \text{ cm}^2.$$

It is usually desirable to subdivide the wire into insulated strands to minimize eddy current losses in the coil winding.

References

1. "Commercial Loading of Telephone Circuits in the Bell System," B. Gherardi, *Trans. A.I.E.E.*, v. 30, 1911, p. 1743.
2. "Development and Application of Loading for Telephone Circuits," T. Shaw and W. Fondiller, *Jour. A.I.E.E.*, v. 45, 1926, p. 253; *B.S.T.J.*, v. 5, 1926, p. 221.
3. "Magnetic Properties of Compressed Powdered Iron," B. Speed and G. W. Elmen, *Trans. A.I.E.E.*, v. 40, 1921, p. 596.
4. "Compressed Powdered Permalloy Manufacture and Magnetic Properties," W. J. Shackelton and I. G. Barber, *Trans. A.I.E.E.*, v. 47, 1928, p. 429.
5. A. F. Bandur, U. S. Patent 1,673,790, June 19, 1928.
6. "Pupinspulen mit Kernen aus Isoperm-Blech oder -Band," H. Jordan, T. Volk and R. Goldschmidt, Europaischer Fernsprechdienst v. 31, 1933, p. 8.
7. "Permalloy, an Alloy of Remarkable Magnetic Properties," H. D. Arnold and G. W. Elmen, *Jour. Frank. Inst.*, v. 195, 1923, p. 621.
8. "Carrier Current Telephony and Telegraphy," E. H. Colpitts and O. B. Blackwell, *Trans. A.I.E.E.*, v. 40, 1921, p. 205.
9. "Radio Extension of the Telephone System to Ships at Sea," H. W. Nichols and L. Espenschied, *Proc. I.R.E.*, v. 11, 1923, p. 193; *B.S.T.J.*, v. 2, 1923, p. 141.
10. "Magnetic Alloys of Iron, Nickel, and Cobalt," G. W. Elmen, *Jour. Frank. Inst.*, v. 207, 1929, p. 583; *B.S.T.J.*, v. 8, 1929, p. 435.
11. "A Survey of Magnetic Materials in Relation to Structure," W. C. Ellis and E. E. Schumacher, *B.S.T.J.*, v. 14, 1935, p. 8.
12. "Statistical Description of the Size Properties of Non-uniform Particulate Substances," T. Hatch and S. Choate, *Jour. Frank. Inst.*, v. 207, 1929, p. 369.
13. "On the Self-induction of Wires," O. Heaviside, *Phil. Mag.*, v. 23, 1887, p. 173.
14. "Magnetostatik der Massekerne," F. Ollendorf, *Arch. f. Elektrotechn*, v. 25, 1931, p. 436.
15. "Iron Powder Compound Cores for Coils," G. W. O. Howe, *Wireless Engineer*, v. 10, 1933, p. 1.
16. "Magnetic Measurements at Low Flux Densities Using the A-C. Bridge," V. E. Legg, *B.S.T.J.*, v. 15, 1936, p. 39.
17. "Harmonic Production in Ferromagnetic Materials at Low Frequencies and Low Flux Densities," E. Peterson, *B.S.T.J.*, v. 7, 1928, p. 762.
18. "Hysteresis Effects with Varying Superposed Magnetizing Forces," W. Fondiller and W. H. Martin, *Trans. A.I.E.E.*, v. 40, 1921, p. 553.
19. "Communication by Carrier in Cable," B. W. Kendall and A. B. Clark, *Elec. Engg.*, v. 52, 1933, p. 477; *B.S.T.J.*, v. 12, 1933, p. 251.
20. "Systems for Wide Band Transmission over Coaxial Lines," L. Espenschied and M. E. Strieby, *Elect. Engg.*, v. 53, 1934, p. 1371; *B.S.T.J.*, v. 13, 1934, p. 654.
21. "The Evolution of the Crystal Wave Filter," O. E. Buckley, *Jour. Appl. Phys.*, v. 8, 1937, p. 40; *Bell Tel. Quarterly*, v. 16, 1937, p. 25.
22. "An Improved Three-Channel Carrier Telephone System," J. T. O'Leary, E. C. Blessing and J. W. Beyer, *B.S.T.J.*, v. 18, 1939, p. 49.

High Accuracy Heterodyne Oscillators

By T. SLONCZEWSKI

The accuracy of a heterodyne oscillator after the low frequency check is made is of the same order of magnitude as that of an ordinary type of oscillator in which circuit elements of the same stability are used. It depends on the constants of the variable frequency oscillator only. This accuracy can be improved by a ratio of 10 to 1 by adding another and higher check frequency. The temperature coefficient of the circuit elements can be kept down to less than 6 parts per million. Scale errors can be reduced to a value comparable with the oscillator accuracy by spreading the scale. A precision oscillator having a frequency range up to 150 kc. and an accuracy of ± 25 cycles including a scale mechanism whereby a large scale spread is obtained on a direct reading scale is described.

Introduction

THE output frequency of a heterodyne oscillator is obtained by modulating the outputs of two oscillators of appreciably higher frequency, one of the oscillators having a fixed frequency, the other being continuously variable over a band width equal to the required output frequency range.

The circuit consists essentially of the two so-called local oscillators, the modulator, where the difference frequency is generated, and an amplifier where the modulator output is raised to the desired level.

The earliest designs of heterodyne oscillator were confined to the audio frequency range, but recently carrier-frequency applications have become more numerous. As the frequency range of the oscillators has increased, their per cent accuracy requirement has increased also. The required frequency accuracy of the oscillator is determined by the maximum slope of the frequency characteristic of the apparatus being measured. If this slope is great, as in the case of a sharply tuned circuit a relatively small displacement of the frequency will result in a large error in the value to be measured. In carrier-frequency systems where the signal is displaced upwards in the frequency scale by modulation, each channel has to meet same crosstalk and transmission requirements independent of its location in the carrier band. Therefore, the maximum slope of the characteristics is independent of the frequency and an oscillator used for measuring purposes has to meet a constant frequency error requirement. In addition the accur-

acy required when expressed in cycles is comparable with that of audio-frequency oscillators so that the percentage accuracy must be much higher.

The advantages of the heterodyne oscillator have made it desirable to study its sources of error to determine whether such an oscillator can be designed to have sufficient accuracy for these applications.

Oscillators With a Single Frequency Check

The frequency of a heterodyne oscillator is given by the expression:

$$F = f' - f, \tag{1}$$

where we will assume f' to be constant and f to be variable and less than f' whence the frequency of the variable frequency oscillator is lowered as the output frequency of the heterodyne oscillator is raised.

The value of F is usually much smaller than either f' or f and relatively small frequency shifts in the local oscillators produced by aging and temperature effects upon the elements of their resonant circuits and changes in vacuum tubes and in the stray capacitances of the circuits produce large relative variations in the output frequency. Usually the stability required of F and the ratio f'/F are so high that it is impracticable to design local oscillators of sufficient stability to meet requirements. Instead, in all heterodyne oscillators an adjustment in the form of a padding condenser in the circuit of the fixed frequency oscillator is used, whereby its frequency is adjusted shortly before the measurement until the oscillator reads correctly at the bottom of its frequency range. The adjustment is made by the zero beat method or by comparison with a low-frequency standard such as a vibrating reed or the 60-cycle power supply.

At the time of the adjustment the frequency of the oscillator is

$$F_o = f' - f_o, \tag{2}$$

where f_o is the value of f at the check frequency F_o. Eliminating f' between (1) and (2) we obtain

$$F = F_o + (f_o - f). \tag{3}$$

The frequency of the variable oscillator may be expressed as

$$f = 1/(2\pi\sqrt{L(C_o + C_a)}), \tag{4}$$

where C_a is the change in the variable air condenser capacitance from the value it has at f_o, and C_o is essentially the value of the fixed con-

denser, usually a good mica unit. L is the inductance of the resonant circuit.

Combining (3) and (4) we get

$$F = F_o + 1/(2\pi\sqrt{LC_o}) - 1/(2\pi\sqrt{L(C_o + C_a)}). \tag{5}$$

The accuracy of the oscillator will depend on the variations in the values of F_o, L, C_o and C_a and is independent of the constants of the fixed frequency oscillator.

By giving increments ΔF_o, ΔC_o, ΔC_a and ΔL to the constants F_o, C_o, C_a and L we obtain after simplifying the expressions

$$\Delta F_{Fo} = \Delta F_o, \tag{6}$$

$$\Delta F_{Co} = -\frac{\Delta C_o}{2C_o} f_o \left[1 - \left(1 + \frac{F_o}{f_o} - \frac{F}{f_o} \right)^3 \right], \tag{7}$$

$$\Delta F_{Ca} = \frac{\Delta C_a}{2C_a} f_o \left[1 + \frac{F_o}{f_o} - \frac{F}{f_o} \right] \left[1 - \left(1 + \frac{F_o}{f_o} - \frac{F}{f_o} \right)^2 \right], \tag{8}$$

$$\Delta F_L = -\frac{\Delta F}{2L} F \left(1 - \frac{F_o}{F} \right) = -\frac{\Delta L}{2L} (F - F_o), \tag{9}$$

giving the corresponding frequency errors ΔF where $f_o = 1/(2\pi\sqrt{LC_o})$ is the variable oscillator frequency at the check frequency F_o.

A variation in F_o will produce an error constant over the whole frequency range. On Fig. 1 the other errors are found plotted in parametric form. To find the error ΔF corresponding to a frequency F the ordinate y corresponding to the value of $x = (F - F_o)/(f_o)$ should be found. Then

$$\Delta F_{Ca} = y_{Ca} f_o \Delta C_a / C_a; \qquad \Delta F_{Co} = y_{Co} \Delta C_o / C_o; \qquad \Delta F_L = y_L f_o \Delta L / L.$$

It is found that F_o can be neglected in all practical cases. The ratio of the ordinate to the abscissa gives the percentage error in frequency caused by a one per cent variation in the element involved.

An examination of the curves shows that they differ only slightly from straight lines which can be interpreted as meaning that the errors are fairly independent of the choice of f_o. This constant should be chosen therefore sufficiently low to require infrequent adjustment at the low-frequency end of the scale. For low values of f_o such that $x > .3$ difficulties in shaping of the air condenser plates and in designing the modulator filter begin to appear. If the errors in an ordinary type of oscillator due to capacitance and inductance variations were plotted on the same set of coordinates the curves would coincide with the line

y_L. This means that if elements of the same accuracy were used, the heterodyne oscillator would be somewhat more accurate. Its total error would be represented by $y_{ca} + y_{co} + y_L$. Since $\Delta C_a/C_a$ and $\Delta C_o/C_o$ will be both positive and of about the same order of magnitude partial compensation will obtain and the error will be of the order of

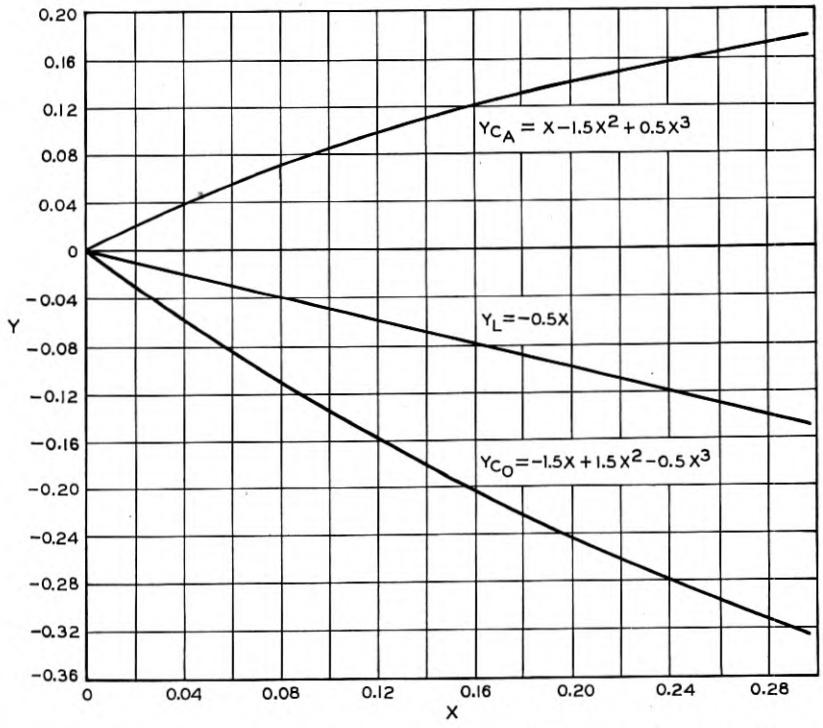

Fig. 1—The frequency errors in a heterodyne oscillator at a frequency $F = xf_0 + F_0$ after the low frequency check had been made can be obtained from the plot as follows: For a variation ΔC_0 in the fixed capacitance C_0, $\Delta F_{c_o} = y_{c_o} \frac{\Delta C_0}{C_0} f_0$; for a variation in the air condenser capacitance C_a, $\Delta F_{c_a} = y_{c_a} \frac{\Delta C_a}{C_a} f_0$; for a variation in the inductance L, $\Delta F_L = y_L \frac{\Delta L}{L} f_0$.

magnitude of ΔF_L. In the case of the ordinary type of oscillator the errors due to the capacitance and inductance variations will be equal and of the same sign so that the error will be of the order of magnitude of $2\Delta F_L$. For audio frequency applications this accuracy has been found to be adequate given sufficient care in the construction of the circuit elements.

Oscillators With a Double Frequency Check

For carrier frequency applications the tolerable error takes a constant value over the entire frequency range and it is found that if a single frequency check is used it is not possible to obtain sufficiently stable elements to maintain the required accuracy at points on the scale removed from the check frequency.

An increase in the accuracy of heterodyne oscillators has been obtained, however, by adding an adjustable condenser to C_o and checking the oscillator at two frequencies, the low frequency F_o and at another, higher, frequency F_s. Adjustment of this condenser by ΔC_o introduces a frequency change—$y_{C_o} f_o \Delta C_o / 2C_o$ adjustable in sign and magnitude and this can be made to cancel the error $\Delta F_{C_a} + \Delta F_L$ for at least one frequency, the check frequency F_s. Obviously if the adjustment is made to correct for variations in C_o no residual error remains. The residual errors which remain after correcting for ΔF_{C_a} and ΔF_L are shown on Fig. 2. The residuals of ΔF_{C_a} and ΔF_L differ from each other

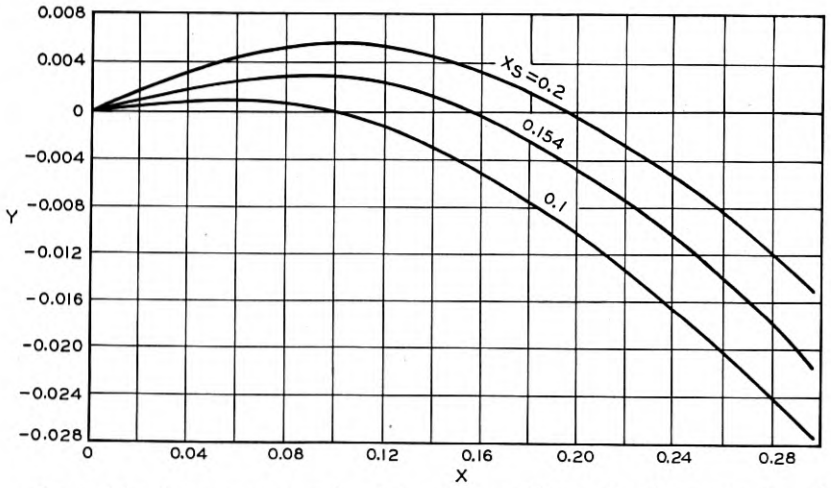

Fig. 2—The frequency errors in a heterodyne oscillator at a frequency $F = xf_0 + F_0$ after the low and high frequency checks had been made can be obtained from the plot as follows: For a variation in the air condenser capacitance C_a, $\Delta F_{C_a} = y \frac{\Delta C_a}{C_a} f_0$; for a variation in the inductance L, $\Delta F_L = y \frac{\Delta L}{L} f_0$.

so little that only one set of curves was drawn. The values of y were obtained by forming the sum $y = Ky_{C_o} + y_{C_a}$ and choosing K so that $y = o$ for $x_s = (F_s - F_o)/f_o$.

For $x_s = .1$ better compensation is obtained at the lower end than at the higher. For a very wide frequency range up to $x = .25$ the best

check frequency would be $x_s = .2$. A good practical limit to x is at 1.9 and here a value of x_s around .15 is best. A further improvement of about 50 per cent could be obtained by choosing a higher value of F_o. When comparing Fig. 2 with Fig. 1 it should be borne in mind that the scale spread for y on Fig. 2 is ten times that of the Fig. 1 which shows that an improvement in accuracy of at least ten to one is obtained by the adjustment. This means, that given two frequency standards F_o and F_s of sufficient accuracy a heterodyne oscillator can be built having a much higher accuracy than an ordinary oscillator having the same frequency range and same quality of circuit elements. This is somewhat contrary to what we are accustomed to think.

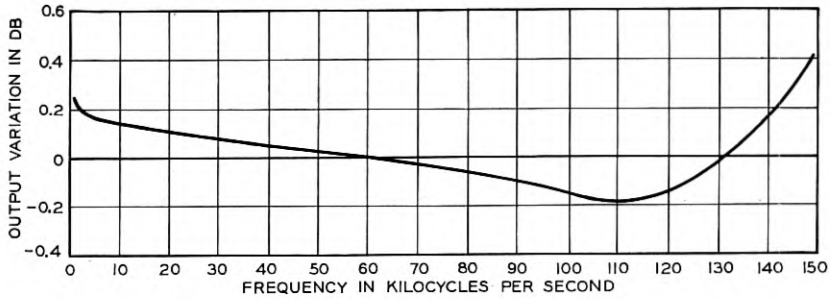

Fig. 3—Frequency-output characteristic.

One detail involved in the procedure of checking the oscillator which permits this high degree of accuracy to be obtained needs elaboration. As C_o is varied during the adjustment by the amount ΔC_o the value of f_o is changed and this destroys the low frequency adjustment at F_o. It is possible to obtain the adjustment by a process of successive approximations but the procedure is tedious. The difficulty can be overcome by the use of a mechanical device, however, as follows. The condenser ΔC_o is ganged to another condenser in the resonant circuit of the fixed oscillator, and the two condensers are so proportioned that the change in the fixed oscillator frequency is equal to the change in f_o as the condenser is adjusted. This makes the low frequency adjustment independent of the high frequency one. The oscillator is just set to the required reading at F_s and ΔC_o is adjusted until the frequency value is correct. Theoretically instead of two condensers two coupled inductometers could have been used to adjust the inductances in the resonant circuits. The net result obtained would have been the same and the mathematical treatment would be like the one given above. Condensers lend themselves better to such construction, however.

Stability of the Constants

Having determined the oscillator errors from the variations in its constants it will be of interest to inquire how large these may be.

When the zero beat method is used the error ΔF_{Fo} will depend on the value of the lowest beat frequency at which the local oscillators can operate. With a reasonable amount of shielding and some precautions in order to avoid mutual inductance in wiring loops it is quite practicable to keep this error below one cycle with local oscillators as high as 200 kc. The beat frequency may be observed on an ammeter placed in the plate circuit of the modulator. When alternating current from the power mains is used as a standard the accuracy is better than one cycle.

There are now available external frequency standards against which the high frequency check could be made which have such high accuracy that the resulting error in the heterodyne oscillator can be entirely neglected. It is desirable, however, to make the oscillator independent of external sources for its adjustment. A convenient checking circuit consists of a quartz crystal which is thrown in with a key across the grids of the output amplifier. At the series resonance frequency of the crystal the loss introduced reduces the output so sharply that the minimum output can be observed within 3 cycles at 100 kc. At any other frequency the error is therefore 30 ppm (parts per million). By using properly cut crystals the temperature variation error is made negligible.

The variations in L are chiefly due to temperature variations. Ordinary potted coils having a large number of layers have temperature coefficients up to 20 parts per million per degree Fahrenheit. The variation is chiefly due to the expansion of the wire.

This error is tolerable in audio frequency oscillators for most purposes. For carrier frequency oscillators unpotted coils having a single layer bank winding wound on a phenol plastic form may be used. Here the lengthwise expansion of the form, which tends to decrease the inductance partly compensates for the expansion of the winding which tends to increase the inductance. Coefficients from 0 to + 6 ppm per °F. are obtained.

The capacitance C_a in commercial air condensers has temperature coefficients of up to 25 ppm per °F. This, again, gives sufficient accuracy for audio frequency oscillators but is not satisfactory for carrier applications. The variations in capacitance with temperature are produced by increase in area of the plates with their expansion which increases by an amount equal to twice the linear coefficient of expansion of the material used. This change is partly compensated by the length-

ening of the air-gaps. When, as usual, several materials are used in the construction, bending of the stator plates due to strains introduced by unequal expansions of the members produce unpredictable changes in capacitance. This is particularly true in the most common construction where the stators are held in place by rods of insulating material. The insulator having a different temperature coefficient of expansion than the plates, the difference in the expansion causes the plates to buckle.

Better stability can be obtained in a condenser built as follows. All parts determining the length of the condenser, including the stator supports and the stator plates are of aluminum. The ends of the stator supports are held in place by insulating bushings of sufficiently small dimensions to make the difference in expansion negligible. The bushings are made of Alsimag, a ceramic material which has a small dielectric constant and coefficient of dielectric constant.

With such a construction, the temperature coefficient of the condenser is equal to twice the temperature coefficient of expansion of the material of which the rotor plates are made minus the temperature coefficient of linear expansion of aluminum determining the length of the air-gaps. One half of the rotor plates are made of invar and one half of aluminum. The average expansion of the area of the rotor plates equals then the temperature coefficient of linear expansion of aluminum and the temperature coefficient of capacitance of the condenser should be equal to the temperature coefficient of air dielectric constant which is about 1 ppm per °F. negative. Measurements show that the temperature coefficient of the condenser varies from $-$ 3 to $+$ 4 ppm per °F., a quite acceptable value. The capacitance change due to a variation in the atmospheric pressure of one inch, a large variation, is 20 ppm.

Temperature coefficients, of paraffined mica condensers, can be adjusted by special manufacturing methods to 10 ppm negative. For the sake of increasing the instantaneous stability the two condensers used in each oscillator are paired within 3 ppm. As mentioned before, no residual error due to ΔC_o remains after the frequency check is made. The low temperature coefficients are desirable only to improve the stability of the oscillator.

By using high Q circuits and suitable corrective reactances, the variations in the frequency due to power line variations may be readily kept smaller than any one of the other errors discussed above.

Scale Errors

A heterodyne oscillator cannot be classified as a purely electrical circuit for it is used to translate a mechanical coordinate, the scale

setting, into an electric coordinate, the output frequency. In planning the oscillator design, therefore, it is necessary to give as much attention to the construction of the scale as to the construction of the circuit elements.

For maximum scale length economy the scale should be so subdivided that a frequency interval equal to the tolerable frequency error ΔF could be read. The scale interval Δl corresponding to this frequency interval, will vary with the measuring conditions. For well illuminated scales on panel mounted equipment to be read conveniently at arm's length an interval Δl of at least .05″ is needed. For portable apparatus, intervals as small as .02″ have been used. With the aid of a vernier it can be brought down to .001″. Scale spreads such that a frequency interval much smaller than ΔF can be read are not only uneconomical but are also objectionable because they encourage the use of the instrument beyond its accuracy limits.

Having chosen Δl and the frequency error ΔF at all points of the scale, the scale shape $l = f(F)$ can be determined by the approximation

$$l = \int_0^F \frac{\Delta l}{\Delta F} dF.$$

As an example, in audio frequency applications the most common form of frequency accuracy desired is that having a constant percentage value $\Delta F/F = \rho$ at the upper part of the scale. At lower frequencies this accuracy is higher than necessary and the requirement is changed to a constant ΔF_o. A smooth shape is obtained by making the transition point F_T at such a frequency that $\Delta F_o/F_T = \rho$. The scale shape is then approximately

$$l = \int_0^F \frac{\Delta l}{\Delta F_o} dF = \frac{\Delta l}{\Delta F_o} F \qquad \text{for} \qquad F < F_T$$

and

$$l = \int_0^{F_T} \frac{\Delta l}{\Delta F_o} dF + \int_{F_T}^F \frac{\Delta l}{\rho F} dF = \frac{\Delta l}{\Delta F_o} F_T + \frac{\Delta l}{\rho} \log_e \frac{F}{F_T}, \quad \text{for} \quad F > F_T$$

The scale of common type of audio frequency oscillator can be spread over a ten inch dial giving a satisfactory accuracy.

For carrier applications, where the spread of any voice band is independent of its position in the frequency range the error function takes the form of a constant and the scale should be linear. Usually the scale lengths involved are much larger than in audio oscillators. To obtain sufficient scale length a precision worm and gear mechanism has to be used to drive the tuning condenser of the heterodyne oscilla-

tor. It gives a scale length of 300 inches, the equivalent of a 5-foot dial and can duplicate settings to better than one part in 10,000.

One detail of construction of such long scales deserves mention. Commercial worm driven air condensers carry on the worm shaft a drum or a dial on which fractions of a revolution of the worm shaft are recorded, while the number of revolutions is recorded on a main dial fixed on the rotor shaft. The effective scale length is then equal to the total displacement of the periphery of the small dial or drum. Using such a construction the oscillator has to be set by consulting a calibration chart where the position of the main dial and of the worm shaft is recorded against the oscillator frequency. Thus one of the most valuable properties of the short scale audio frequency oscillator, its direct reading, is lost.

To remedy this situation a special scale mechanism has been developed for carrier frequency oscillators which combines great scale length with good spread and compactness. It consists of a long motion picture film strip engaged by a two inch film sprocket mounted in place of the conventional drum on the worm shaft. The rotation of the shaft determines the displacement of the film against an index which reads the frequency directly in kilocycles. The loose ends of the strip are wound up on two spools interconnected by a spring mechanism which takes up the slack. The whole mechanism is confined in a space about 4″ by 5″ by 5″ accommodating a scale length up to 450 inches with a scale spread corresponding to $\Delta l = .05''$.

Specific Application

An example of application of these methods in the design of a heterodyne oscillator is furnished by an oscillator built for use in connection with the installation and maintenance of broad band transmission systems. It is shown on Fig. 4.

It has a frequency range of from 1 to 150 kc. Its variable frequency oscillator covers a range of from 500 to 650 kc. This was chosen as low as possible to obtain good instantaneous stability, but high enough not to introduce difficulties in designing the filter following the modulator. The capacitance of the air condenser is about 800 $\mu\mu$f. From the circuit design standpoint a larger capacitance would be desirable, but for the stability required the overall size of the condenser sets an upper limit to the capacitance. The frequency and air condenser capacitance values determine the value of the fixed condenser at 1000 mmf and the coil inductance at 50 microhenries. The fixed oscillator is similar to the variable oscillator except for the omission of the variable air condenser.

The frequency setting is recorded on a 300-inch film scale such as described above. This gives a spread of two inches per kilocycle. With the 50-cycle divisions marked directly the mechanism can be readily set to an accuracy better than 25 cycles. The visibility of the scale is greatly enhanced by a pilot lamp placed in back of the scale

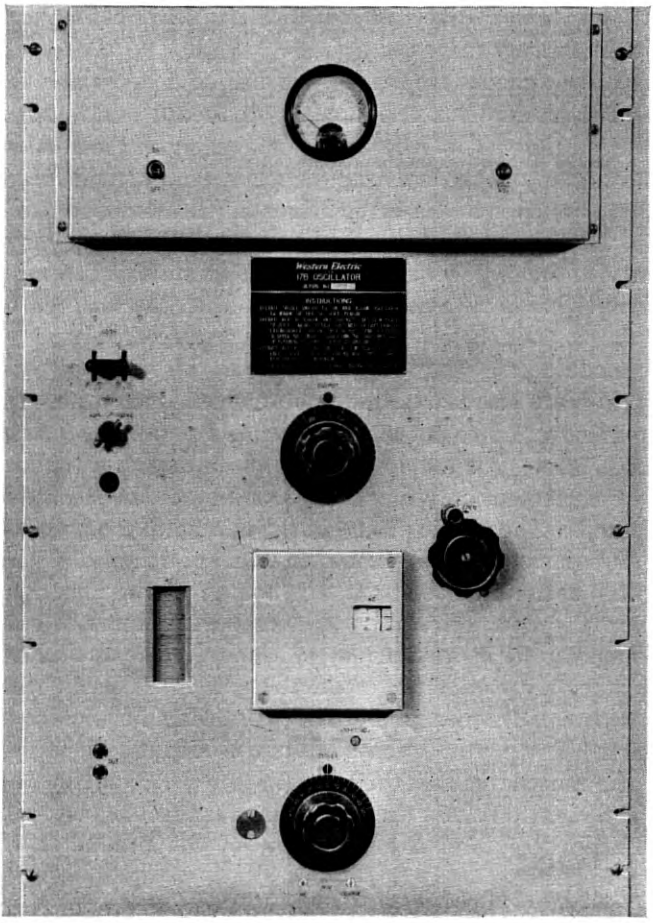

Fig. 4—Front view of the oscillator.

window with an intervening opal glass. A crank on the front of the panel is used to set the oscillator, the range being covered in 47 revolutions. When changing the frequency setting even at a moderate rate the speed with which the film moves prevents the operator from observing the frequency setting. To make the adjustment more convenient, a coarse scale is recorded on a dial which can be read easily to one

kilocycle while the mechanism is in motion. It can be seen under the hood in the center of the panel.

Below the coarse frequency dial is seen a small dial connected to a variable condenser which permits the operator to vary the frequency of the oscillator up to ± 50 cycles from the frequency to which it is set and to read the frequency change with an accuracy of about 3 cycles. This feature is found to be useful in locating peaks of frequency characteristics of sharply resonant circuits.

The frequency checks are made by operating a key which throws the oscillator output across a telephone switchboard lamp and a 100 kc crystal across the grid of the output stage. For the low frequency check another key superposes the 60 cycles power main frequency on the oscillator output and a screwdriver adjustment operating a condenser in the fixed oscillator adjusts the oscillator frequency to synchronism with the scale set at 60 cycles. For the high frequency check a minimum signal is obtained on the lamp with the scale set to 100 kc by adjusting a padding condenser in the variable oscillator.

In the modulator a pentode type vacuum tube is used, which has a control grid-plate current characteristic which over nearly the entire region from zero bias to cut-off approaches a parabola so closely, that where modulation products lower than 40 db down on the useful output can be neglected, only first and second modulation products need be considered. The bias is placed in the middle of the parabolic range and the two input signals are adjusted to equality and to a value covering the entire parabolic range. This gives the maximum useful modulation output necessitating the smallest amount of gain in the output stage at little sacrifice in efficiency. The modulator being parabolic, the only products of modulation other than the useful output are the two high frequency input signals, their harmonics and sum frequencies. These are eliminated from the output by inserting a filter between the modulator and the output stage. Advantage is taken of phase discrimination since the circuit is arranged in push-pull to decrease the filter requirements for some of the products, which are generated in phase.

The plate supply is obtained from a rectifier operating on the 60-cycle main supply. It is provided with a vacuum tube regulator circuit which keeps the plate and screen voltages constant over a ± 5 volt variation of the power line voltage. The output control is obtained by means of a potentiometer in the output amplifier input. Two output impedances, 600 and 135 ohms, may be selected by operating a key.

The apparatus is mounted on a standard 19-inch panel 28 inches high. The bottom, the coolest part, is occupied by the oscillators; the middle

by the modulator and amplifier; and the top by the power pack. Perforations in the oscillator cover provide ventilation to reduce warming-up effects. A close-up giving the details of the scale mechanism and the shielding is shown on Fig. 5.

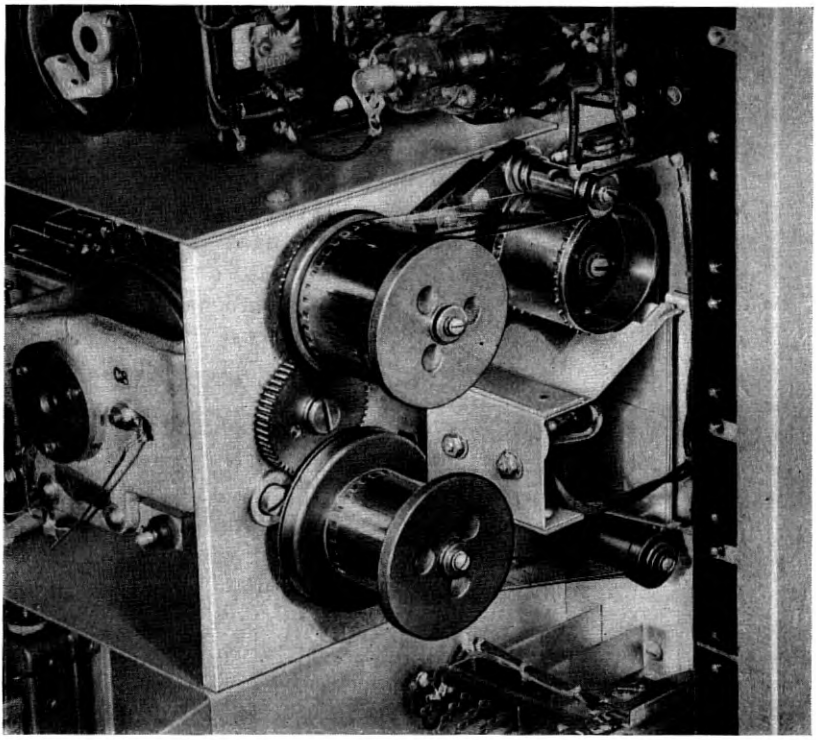

Fig. 5—Details of the scale mechanism.

Tests on the oscillator show that the overall frequency accuracy throughout its range can be maintained to ± 25 cycles. The harmonics are down 40 db from the fundamental at 100-milliwatt output. With the full output of one watt the harmonics are 30 db down. The total output variation with frequency is shown on Fig. 3.

This oscillator has found a wide range of applications as an accurate source of frequency in the communications field.

List of Symbols

F output frequency of the heterodyne oscillator

F_o standard frequency used to check the oscillator at the low end of the scale

F_s standard frequency used to check the oscillator at the high end of the scale

F_T frequency at which the scale changes from linear to logarithmic

l length of the scale interval from 0 to F

f frequency of the variable oscillator

f' frequency of the fixed oscillator

f_o frequency of the variable oscillator at the setting $F = F_o$

L inductance in the resonant circuit of variable oscillator

C total capacitance in the resonant circuit of variable oscillator

C_o total capacitance in the resonant circuit of variable oscillator when set to $F = F_o$

$C_a = C - C_o$ capacitance change in the air condenser

ΔF_o variation in the standard frequency F_o

ΔC_a variation in C_a

ΔC_o variation in C_o

ΔL variation in L

Δl smallest readable scale interval

ρ relative frequency error $\Delta F/F$

ΔF error in F

ΔF_{Fo} error in F caused by ΔF_o

ΔF_{Co} error in F caused by ΔC_o

ΔF_{Ca} error in F caused by ΔC_a

References

"Beat-Frequency Oscillators," M. S. Mead, Jr. *G. E. Rev.*, v. 32, pp. 521–529, Oct., 1929.

"Beat-Frequency Oscillator," M. F. Cooper and L. G. Page, *Wireless Engineer*, v. 10, pp. 469–476, Sept., 1933.

"Precision Heterodyne Oscillators," D. A. Bell, *Wireless Engineer*, v. 11, p. 308, June, 1934.

"Precision Heterodyne Oscillators," W. H. F. Griffiths, *Wireless Engineer*, v. 11, pp. 234–244, May, 1934.

"A Precision Heterodyne Oscillator," L. E. Ryall, *P. O. E. E., Jl.*, v. 27, pp. 213–221, Oct., 1934.

"The Beat-Frequency Oscillator," M. Slaffer, *Wireless Engineer*, v. 11, pp. 25–26, Jan., 1934. (Letter to the Editor.)

"Further Notes on Precision Heterodyne Oscillators," W. H. F. Griffiths, *Wireless Engineer*, v. 12, pp. 357–362, July, 1935.

"Beat Frequency Oscillators," A. W. Barber, *Radio Engg.*, v. 16, pp. 13–16, July, 1936.

"A Beat-Frequency Oscillator," L. B. Hallman, *Commun. & Broadcast Engg.*, v. 3, pp. 10–13, May, 1936.

Relations Between Attenuation and Phase in Feedback Amplifier Design

By H. W. BODE

Introduction

THE engineer who embarks upon the design of a feedback amplifier must be a creature of mixed emotions. On the one hand, he can rejoice in the improvements in the characteristics of the structure which feedback promises to secure him.[1] On the other hand, he knows that unless he can finally adjust the phase and attenuation characteristics around the feedback loop so the amplifier will not spontaneously burst into uncontrollable singing, none of these advantages can actually be realized. The emotional situation is much like that of an impecunious young man who has impetuously invited the lady of his heart to see a play, unmindful, for the moment, of the limitations of the $2.65 in his pockets. The rapturous comments of the girl on the way to the theater would be very pleasant if they were not shadowed by his private speculation about the cost of the tickets.

In many designs, particularly those requiring only moderate amounts of feedback, the bogy of instability turns out not to be serious after all. In others, however, the situation is like that of the young man who has just arrived at the box office and finds that his worst fears are realized. But the young man at least knows where he stands. The engineer's experience is more tantalizing. In typical designs the loop characteristic is always satisfactory—except for one little point. When the engineer changes the circuit to correct that point, however, difficulties appear somewhere else, and so on ad infinitum. The solution is always just around the corner.

Although the engineer absorbed in chasing this rainbow may not realize it, such an experience is almost as strong an indication of the existence of some fundamental physical limitation as the census which the young man takes of his pockets. It reminds one of the experience of the inventor of a perpetual motion machine. The perpetual motion machine, likewise, always works—except for one little factor. Evidently, this sort of frustration and lost motion is inevitable in

[1] A general acquaintance with feedback circuits and the uses of feedback is assumed in this paper. As a broad reference, see H. S. Black, "Stabilized Feedback Amplifiers," *B. S. T. J.*, January, 1934.

feedback amplifier design as long as the problem is attacked blindly. To avoid it, we must have some way of determining in advance when we are either attempting something which is beyond our resources, like the young man on the way to the theater, or something which is literally impossible, like the perpetual motion enthusiast.

This paper is written to call attention to several simple relations between the gain around an amplifier loop, and the phase change around the loop, which impose limits to what can and cannot be done in a feedback design. The relations are mathematical laws, which in their sphere have the same inviolable character as the physical law which forbids the building of a perpetual motion machine. They show that the attempt to build amplifiers with certain types of loop characteristics *must* fail. They permit other types of characteristic, but only at the cost of certain consequences which can be calculated. In particular, they show that the loop gain cannot be reduced too abruptly outside the frequency range which is to be transmitted if we wish to secure an unconditionally stable amplifier. It is necessary to allow at least a certain minimum interval before the loop gain can be reduced to zero.

The question of the rate at which the loop gain is reduced is an important one, because it measures the actual magnitude of the problem confronting both the designer and the manufacturer of the feedback structure. Until the loop gain is zero, the amplifier will sing unless the loop phase shift is of a prescribed type. The cutoff interval as well as the useful transmission band is therefore a region in which the characteristics of the apparatus must be controlled. The interval represents, in engineering terms, the price of the ticket.

The price turns out to be surprisingly high. It can be minimized by accepting an amplifier which is only conditionally stable.[2] For the customary absolutely stable amplifier, with ordinary margins against singing, however, the price in terms of cutoff interval is roughly one octave for each ten db of feedback in the useful band. In practice, an additional allowance of an octave or so, which can perhaps be regarded as the tip to the hat check girl, must be made to insure that the amplifier, having once cut off, will stay put. Thus in an amplifier with 30 db feedback, the frequency interval over which effective control of the loop transmission characteristics is necessary is at least four octaves, or sixteen times, broader than the useful band. If we raise the feedback to 60 db, the effective range must be more than a hundred times the useful range. If the useful band is itself large these factors

[2] Definitions of conditionally and unconditionally stable amplifiers are given on page 432.

may lead to enormous effective ranges. For example, in a 4 megacycle amplifier they indicate an effective range of about 60 megacycles for 30 db feedback, or of more than 400 megacycles if the feedback is 60 db.

The general engineering implications of this result are obvious. It evidently places a burden upon the designer far in excess of that which one might anticipate from a consideration of the useful band alone. In fact, if the required total range exceeds the band over which effective control of the amplifier loop characteristics is physically possible, because of parasitic effects, he is helpless. Like the young man, he simply can't pay for his ticket. The manufacturer, who must construct and test the apparatus to realize a prescribed characteristic over such wide bands, has perhaps a still more difficult problem. Unfortunately, the situation appears to be an inevitable one. The mathematical laws are inexorable.

Aside from sounding this warning, the relations between loop gain and loop phase can also be used to establish a definite method of design. The method depends upon the development of overall loop characteristics which give the optimum result, in a certain sense, consistent with the general laws. This reduces actual design procedure to the simulation of these characteristics by processes which are essentially equivalent to routine equalizer design. The laws may also be used to show how the characteristics should be modified when the cutoff interval approaches the limiting band width established by the parasitic elements of the circuit, and to determine how the maximum realizable feedback in any given situation can be calculated. These methods are developed at some length in the writer's U. S. Patent No. 2,123,178 and are explained in somewhat briefer terms here.

Relations Between Attenuation and Phase in Physical Networks [3]

The amplifier design theory advanced here depends upon a study of the transmission around the feedback loop in terms of a number of general laws relating the attenuation and phase characteristics of physical networks. In attacking this problem an immediate difficulty presents itself. It is apparent that no entirely definite and universal

[3] Network literature includes a long list of relations between attenuation and phase discovered by a variety of authors. They are derived typically from a Fourier analysis of the transient response of assumed structures and are frequently ambiguous, because of failure to recognize the minimum phase shift condition. No attempt is made to review this work here, although special mention should be made of Y. W. Lee's paper in the *Journal for Mathematics and Physics* for June, 1932. The proof of the relations given in the present paper depends upon a contour integration in the complex frequency plane and can be understood from the disclosure in the patent referred to previously.

relation between the attenuation and the phase shift of a physical structure can exist. For example, we can always change the phase shift of a circuit without affecting its loss by adding either an ideal transmission line or an all-pass section. Any attenuation characteristic can thus correspond to a vast variety of phase characteristics.

For the purposes of amplifier design this ambiguity is fortunately unimportant. While no unique relation between attenuation and phase can be stated for a general circuit, a unique relation does exist between any given loss characteristic and the *minimum* phase shift which must be associated with it. In other words, we can always add a line or all-pass network to the circuit but we can never subtract such a structure, unless, of course, it happens to be part of the circuit originally. If the circuit includes no surplus lines or all-pass sections, it will have at every frequency the least phase shift (algebraically) which can be obtained from any physical structure having the given attenuation characteristic. The least condition, since it is the most favorable one, is, of course, of particular interest in feedback amplifier design.

For the sake of precision it may be desirable to restate the situations in which this minimum condition fails to occur. The first situation is found when the circuit includes an all-pass network either as an individual structure or as a portion of a network which can be replaced by an all-pass section in combination with some other physical structure.[4] The second situation is found when the circuit includes a transmission line. The third situation occurs when the frequency is so high that the tubes, network elements and wiring cannot be considered to obey a lumped constant analysis. This situation may be found, for example, at frequencies for which the transit time of the tubes is important or for which the distance around the feedback loop is an appreciable part of a wave-length. The third situation is, in many respects, substantially the same as the second, but it is mentioned separately here as a matter of emphasis. Since the effective band of a feedback amplifier is much greater than its useful band, as the introduction pointed out, the considerations it reflects may be worth taking into account even when they would be trivial in the useful band alone.

It will be assumed here that none of these exceptional situations is found. For the minimum phase condition, then, it is possible to derive

[4] Analytically this condition can be stated as follows: Let it be supposed that the transmission takes place between mesh 1 and mesh 2. The circuit will include an all-pass network, explicit or concealed, if any of the roots of the minor Δ_{12} of the principal circuit determinant lie below the real axis in the complex frequency plane. This can happen in bridge configurations, but not in series-shunt configurations, so that all ladder networks are automatically of minimum phase type.

a large number of relations between the attenuation and phase characteristics of a physical network. One of the simplest is

$$\int_{-\infty}^{\infty} B du = \frac{\pi}{2}(A_\infty - A_0), \tag{1}$$

where u represents $\log f/f_0$, f_0 being an arbitrary reference frequency, B is the phase shift in radians, and A_0 and A_∞ are the attenuations in nepers at zero and infinite frequency, respectively. The theorem states, in effect, that the total area under the phase characteristic plotted on a logarithmic frequency scale depends only upon the difference between the attenuations at zero and infinite frequency, and not upon the course of the attenuation between these limits. Nor does it depend upon the physical configuration of the network unless a non-minimum phase structure is chosen, in which case the area is necessarily increased. The equality of phase areas for attenuation characteristics of different types is illustrated by the sketches of Fig. 1.

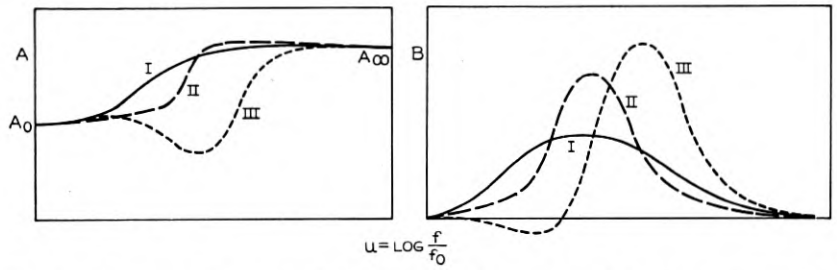

Fig. 1—Diagram to illustrate relation between phase area and change in attenuation.

The significance of the phase area relation for feedback amplifier design can be understood by supposing that the practical transmission range of the amplifier extends from zero to some given finite frequency. The quantity $A_0 - A_\infty$ can then be identified with the change in gain around the feedback loop required to secure a cut-off. Associated with it must be a certain definite phase area. If we suppose that the maximum phase shift at any frequency is limited to some rather low value the total area must be spread out over a proportionately broad interval on the frequency scale. This must correspond roughly to the cut-off region, although the possibility that some of the area may be found above or below the cut-off range prevents us from determining the necessary interval with precision.

A more detailed statement of the relationship between phase shift and change in attenuation can be obtained by turning to a second

theorem. It reads as follows:

$$B(f_c) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{dA}{du} \log \coth \frac{|u|}{2}\, du, \tag{2}$$

where $B(f_c)$ represents the phase shift at any arbitrarily chosen frequency f_c and $u = \log f/f_c$. This equation, like (1), holds only for the minimum phase shift case.

Although equation (2) is somewhat more complicated than its predecessor, it lends itself to an equally simple physical interpretation. It is clear, to begin with, that the equation implies broadly that the

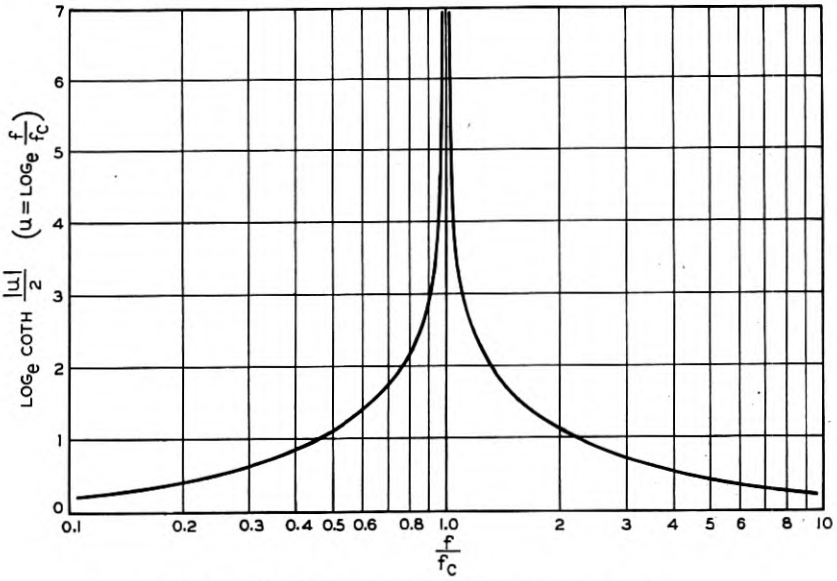

Fig. 2—Weighting function in loss-phase formula.

phase shift at any frequency is proportional to the derivative of the attenuation on a logarithmic frequency scale. For example, if dA/du is doubled B will also be doubled. The phase shift at any particular frequency, however, does not depend upon the derivative of attenuation at that frequency alone, but upon the derivative at all frequencies, since it involves a summing up, or integration, of contributions from the complete frequency spectrum. Finally, we notice that the contributions to the total phase shift from the various portions of the frequency spectrum do not add up equally, but rather in accordance with the function log coth $|u|/2$. This quantity, therefore, acts as a weighting function. It is plotted in Fig. 2. As we might expect physically

it is much larger near the point $u = 0$ than it is in other regions. We can, therefore, conclude that while the derivative of attenuation at all frequencies enters into the phase shift at any particular frequency $f = f_c$ the derivative in the neighborhood of f_c is relatively much more important than the derivative in remote parts of the spectrum.

As an illustration of (2), let it be supposed that $A = ku$, which corresponds to an attenuation having a constant slope of $6\ k$ db per octave. The associated phase shift is easily evaluated. It turns out, as we might expect, to be constant, and is equal numerically to $k\pi/2$ radians. This is illustrated by Fig. 3. As a second example, we may consider

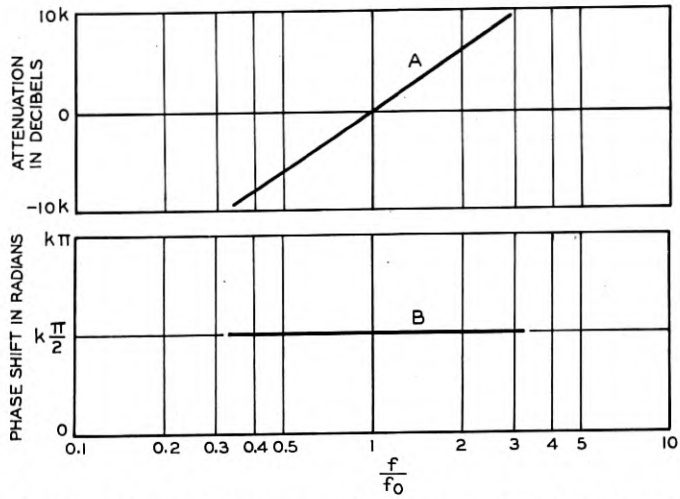

Fig. 3—Phase characteristic corresponding to a constant slope attenuation.

a discontinuous attenuation characteristic such as that shown in Fig. 4. The associated phase characteristic, also shown in Fig. 4, is proportional to the weighting function of Fig. 2.

The final example is shown by Fig. 5. It consists of an attenuation characteristic which is constant below a specified frequency f_b and has a constant slope of $6\ k$ db per octave above f_b. The associated phase characteristic is symmetrical about the transition point between the two ranges. At sufficiently high frequencies, the phase shift approaches the limiting $k\pi/2$ radians which would be realized if the constant slope were maintained over the complete spectrum. At low frequencies the phase shift is substantially proportional to frequency and is given by the equation

$$B = \frac{2k}{\pi}\frac{f}{f_b}. \tag{3}$$

Solutions developed in this way can be added together, since it is apparent from the general relation upon which they are based that the phase characteristic corresponding to the sum of two attenuation

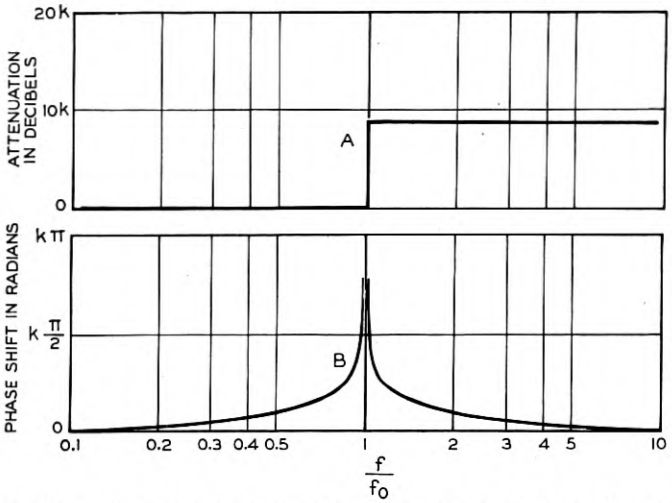

Fig. 4—Phase characteristic corresponding to a discontinuity in attenuation.

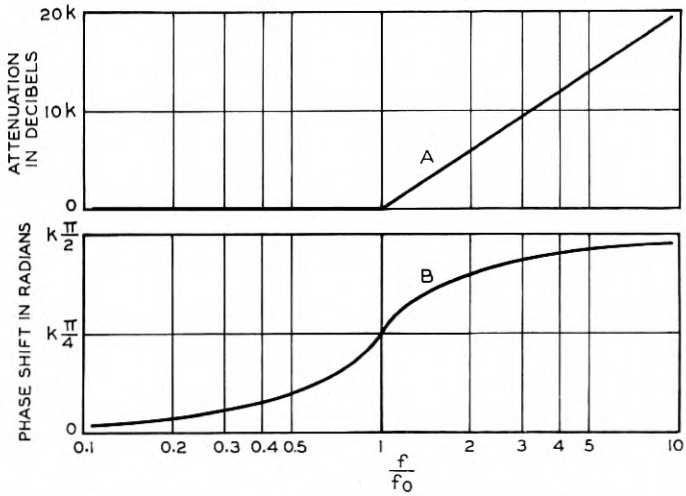

Fig. 5—Phase characteristic corresponding to an attenuation which is constant below a prescribed frequency and has a constant slope above it.

characteristics will be equal to the sum of the phase characteristics corresponding to the two attenuation characteristics separately. We can therefore combine elementary solutions to secure more complicated

characteristics. An example is furnished by Fig. 6, which is built up from three solutions of the type shown by Fig. 5. By proceeding sufficiently far in this way, an approximate computation of the phase characteristic associated with almost any attenuation characteristic can be made, without the labor of actually performing the integration in (2).

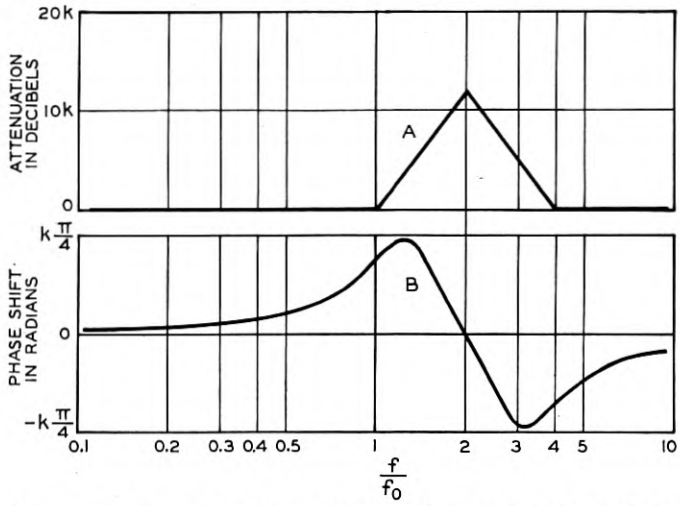

Fig. 6—Diagram to illustrate addition of elementary attenuation and phase characteristics to produce more elaborate solutions of the loss-phase formula.

Equations (1) and (2) are the most satisfactory expressions to use in studying the relation between loss and phase in a broad physical sense. The mechanics of constructing detailed loop cut-off characteristics, however, are simplified by the inclusion of one other, somewhat more complicated, formula. It appears as

$$\int_0^{f_0} \frac{A\,df}{\sqrt{f_0^2 - f^2}\,(f^2 - f_c^2)} + \int_{f_0}^{\infty} \frac{B\,df}{\sqrt{f^2 - f_0^2}\,(f^2 - f_c^2)}$$

$$= \frac{\pi}{2 f_c \sqrt{f_0^2 - f_c^2}}, \qquad f_c < f_0$$

$$= -\frac{\pi}{2 f_c \sqrt{f_c^2 - f_0^2}}, \qquad f_c > f_0, \quad (4)$$

where f_0 is some arbitrarily chosen frequency and the other symbols have their previous significance.

The meaning of (4) can be understood if it is recalled that (2) implies that the minimum phase shift at any frequency can be computed if the

attenuation is prescribed at all frequencies. In the same way (4) shows how the complete attenuation and phase characteristics can be determined if we begin by prescribing the attenuation below f_0 and the phase shift above f_0. Since f_0 can be chosen arbitrarily large or small this is evidently a more general formula than either (1) or (2), while it can itself be generalized, by the introduction of additional irrational factors, to provide for more elaborate patterns of bands in which A and B are specified alternately.

As an example of this formula, let it be assumed that $A = K$ for $f < f_0$ and that $B = k\pi/2$ for $f > f_0$. These are shown by the solid lines in Fig. 7. Substitution in (4) gives the A and B characteristics in the rest of the spectrum as

$$\begin{aligned} B &= k \sin^{-1} \frac{f}{f_0}, && f < f_0 \\ A &= K + k \log \left[\sqrt{\frac{f^2}{f_0^2} - 1} + \frac{f}{f_0} \right], && f > f_0. \end{aligned} \tag{5}$$

These are indicated by broken lines in Fig. 7. In this particularly

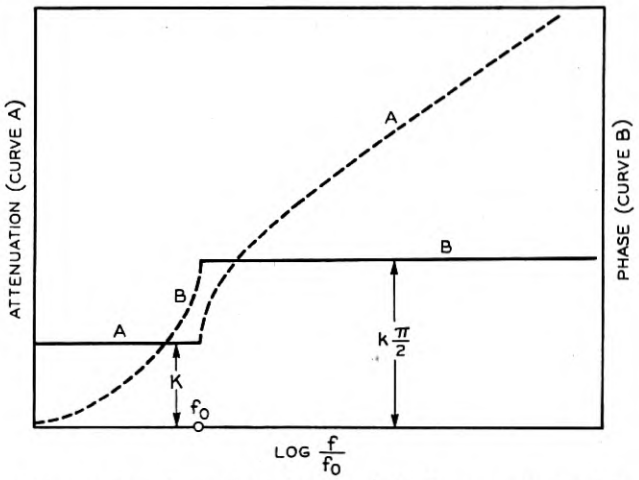

Fig. 7—Construction of complete characteristics from an attenuation characteristic specified below a certain frequency and a phase characteristic above it. The solid lines represent the specified attenuation and phase characteristics, and the broken lines their computed extensions to the rest of the spectrum.

simple case all four fragments can be combined into the single analytic formula

$$A + iB = K + k \log \left[\sqrt{1 - \frac{f^2}{f_0^2}} + i \frac{f}{f_0} \right]. \tag{6}$$

This expression will be used as the fundamental formula for the loop cut-off characteristic in the next section.

Overall Feedback Loop Characteristics

The survey just concluded shows what combinations of attenuation and phase characteristics are physically possible. We have next to determine which of the available combinations is to be regarded as representing the transmission around the overall feedback loop. The choice will naturally depend somewhat upon exactly what we assume that the amplifier ought to do, but with any given set of assumptions it is possible, at least in theory, to determine what combination is most appropriate.

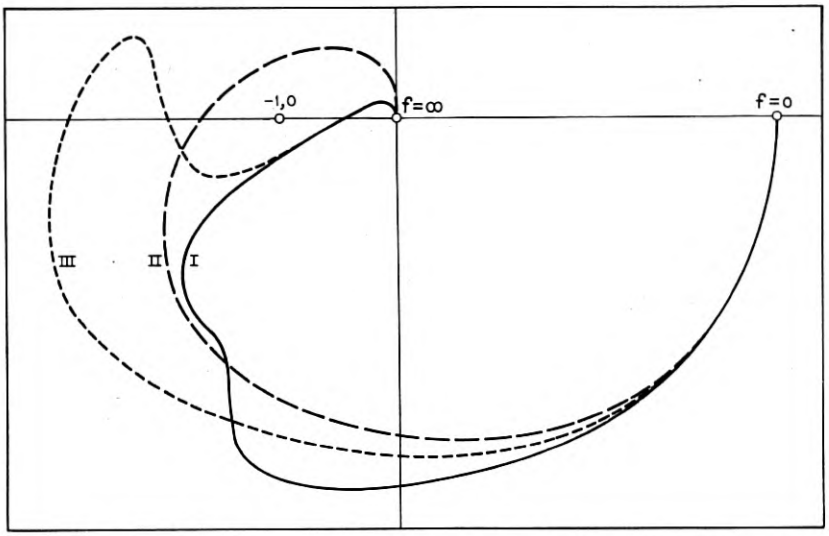

Fig. 8—Nyquist stability diagrams for various amplifiers. Curve I represents "absolute" stability, Curve II instability, and Curve III "conditional" stability. In accordance with the convention used in this paper the diagram is rotated through 180° from its normal position so that the critical point occurs at $-1, 0$ rather than $+1, 0$.

The situation is conveniently investigated by means of the Nyquist stability diagram [5] illustrated by Fig. 8. The diagram gives the path

[5] *Bell System Technical Journal*, July, 1932. See also Peterson, Kreer, and Ware, *Bell System Technical Journal*, October, 1934. The Nyquist diagrams in the present paper are rotated through 180° from the positions in which they are usually drawn, turning the diagrams in reality into plots of $-\mu\beta$. In a normal amplifier there is one net phase reversal due to the tubes in addition to any phase shifts chargeable directly to the passive networks in the circuit. The rotation of the diagram allows this phase reversal to be ignored, so that the phase shifts actually shown are the same as those which are directly of design interest.

traced by the vector representing the transmission around the feedback loop as the frequency is assigned all possible real values. In accordance with Nyquist's results a path such as II, which encircles the point $-1, 0$, indicates an unstable circuit and must be avoided. A stable amplifier is obtained if the path resembles either I or III, neither of which encircles $-1, 0$. The stability represented by Curve III, however, is only "Nyquist" or "conditional." The path will enclose the critical point if it is merely reduced in scale, which may correspond physically to a reduction in tube gain. Thus the circuit may sing when the tubes begin to lose their gain because of age, and it may also sing, instead of behaving as it should, when the tube gain increases from zero as power is first applied to the circuit. Because of these possibilities conditional stability is usually regarded as undesirable and the present discussion will consequently be restricted to "absolutely" or "unconditionally" stable amplifiers having Nyquist diagrams of the type resembling Curve I.

The condition that the amplifier be absolutely stable is evidently that the loop phase shift should not exceed 180° until the gain around the loop has been reduced to zero or less. A theoretical characteristic which just met this requirement, however, would be unsatisfactory, since it is inevitable that the limiting phase would be exceeded in fact by minor deviations introduced either in the detailed design of the amplifier or in its construction. It will therefore be assumed that the limiting phase is taken as 180° less some definite margin. This is illustrated by Fig. 9, the phase margin being indicated as $y\pi$ radians. At frequencies remote from the band it is physically impossible, in most circuits, to restrict the phase within these limits. As a supplement, therefore, it will be assumed that larger phase shifts are permissible if the loop gain is x db below zero. This is illustrated by the broken circular arc in Fig. 9. A theoretical loop characteristic meeting both requirements will be developed for an amplifier transmitting between zero and some prescribed limiting frequency with a constant feedback, and cutting off thereafter as rapidly as possible. This basic characteristic can be adapted to amplifiers with varying feedback in the useful range or with useful ranges lying in other parts of the spectrum by comparatively simple modifications which are described at a later point. It is, of course, contemplated that the gain and phase margins x and y will be chosen arbitrarily in advance. If we choose large values we can permit correspondingly large tolerances in the detailed design and construction of the apparatus without risk of instability. It turns out, however, that with a prescribed width of cutoff interval the amount of feedback which can be realized in the

useful range is decreased as the assumed margins are increased, so that it is generally desirable to choose as small margins as is safe.

The essential feature in this situation is the requirement that the diminution of the loop gain in the cutoff region should not be accompanied by a phase shift exceeding some prescribed amount. In view of the close connection between phase shift and the slope of the attenuation characteristic evidenced by (2) this evidently demands that the amplifier should cut off, on the whole, at a well defined rate which is not too fast. As a first approximation, in fact, we can choose the cutoff

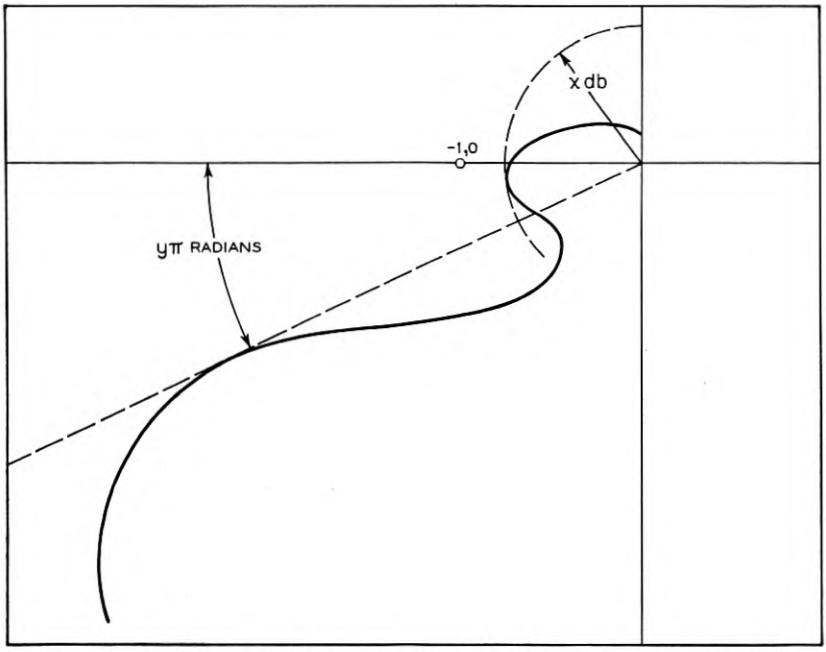

Fig. 9—Diagram to illustrate definitions of phase and gain margins for the feedback loop.

characteristic as an exactly constant slope from the edge of the useful band outward. Such a characteristic has already been illustrated by Fig. 5 and is shown, replotted,[6] by the broken lines in Fig. 10. If we choose the parameter corresponding to k in Fig. 5 as 2 the cutoff rate is 12 db per octave and the phase shift is substantially 180° at high frequencies. This choice thus leads to zero phase margin. By choosing a somewhat smaller k on the other hand, we can provide a definite

[6] To prevent confusion it should be noticed that the general attenuation-phase diagrams are plotted in terms of relative loss while loop cutoff characteristics, here and at later points, are plotted in terms of relative gain.

margin against singing, at the cost of a less rapid cutoff. For example, if we choose $k = 1.5$ the limiting phase shift in the $\mu\beta$ loop becomes 135°, which provides a margin of 45° against instability, while the rate of cutoff is reduced to 9 db per octave. The value $k = 1.67$, which corresponds to a cutoff rate of 10 db per octave and a phase margin of 30°, has been chosen for illustrative purposes in preparing Fig. 10. The loss margin depends upon considerations which will appear at a later point.

Although characteristics of the type shown by Fig. 5 are reasonably satisfactory as amplifier cutoffs they evidently provide a greater phase

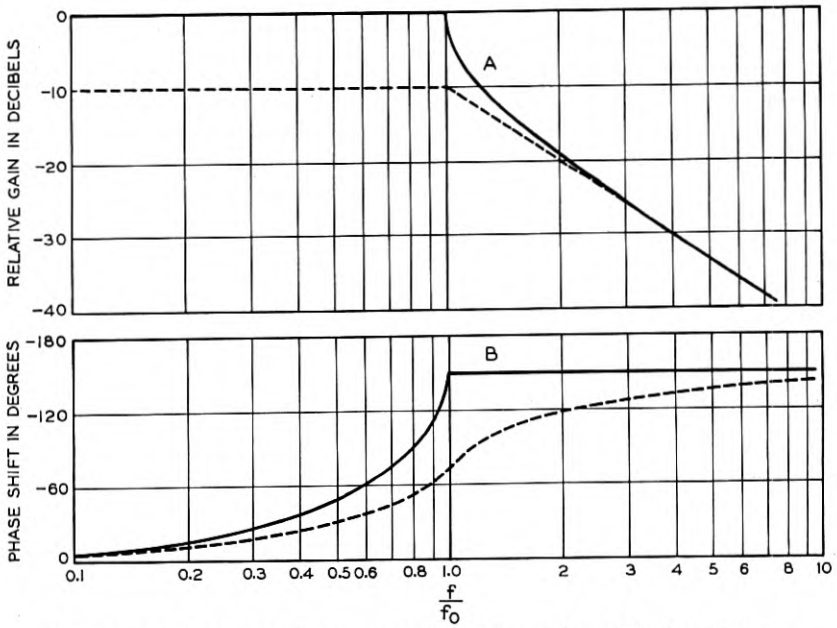

Fig. 10—Ideal loop cutoff characteristics. Drawn for a 30° phase margin.

margin against instability in the region just beyond the useful band than they do at high frequencies. In virtue of the phase area law this must be inefficient if, as is supposed here, the optimum characteristic is one which would provide a constant margin throughout the cutoff interval. The relation between the phase and the slope of the attenuation suggests that a constant phase margin can be obtained by increasing the slope of the cutoff characteristic near the edge of the band, leaving its slope at more remote frequencies unchanged, as shown by the solid lines in Fig. 10. The exact expression for the required curve can be found from (6), where the problem of determining such a characteristic appeared as an example of the use of the general formula (4).

At high frequencies the new phase and attenuation characteristics merge with those obtained from the preceding straight line cutoff, as Fig. 10 indicates. In this region the relation between phase margin and cutoff slope is fixed by the k in the equation (6) in the manner already described for the more elementary cutoff. At low frequencies, however, the increased slope near the edge of the band permits 6 k db more feedback.

It is worth while to pause here to consider what may be said, on the basis of these characteristics, concerning the breadth of cutoff interval required for a given feedback, or the "price of the ticket," as it was expressed in the introduction. If we adopt the straight line cutoff and assume the k used in Fig. 10 the interval between the edge of the useful band and the intersection of the characteristic with the zero gain axis is evidently exactly 1 octave for each 10 db of low frequency feedback. The increased efficiency of the solid line characteristic saves one octave of this total if the feedback is reasonably large to begin with. This apparently leads to a net interval one or two octaves narrower than the estimates made in the introduction. The additional interval is required to bridge the gap between a purely mathematical formula such as (6), which implies that the loop characteristics follow a prescribed law up to indefinitely high frequencies, and a physical amplifier, whose ultimate loop characteristics vary in some uncontrollable way. This will be discussed later. It is evident, of course, that the cutoff interval will depend slightly upon the margins assumed. For example, if the phase margin is allowed to vanish the cutoff rate can be increased from 10 to 12 db per octave. This, however, is not sufficient to affect the order of magnitude of the result. Since the diminished margin is accompanied by a corresponding increase in the precision with which the apparatus must be manufactured such an economy is, in fact, a Pyrrhic victory unless it is dictated by some such compelling consideration as that described in the next section.

Maximum Obtainable Feedback

A particularly interesting consequence of the relation between feedback and cutoff interval is the fact that it shows why we cannot obtain unconditionally stable amplifiers with as much feedback as we please. So far as the purely theoretical construction of curves such as those in Fig. 10 is concerned, there is clearly no limit to the feedback which can be postulated. As the feedback is increased, however, the cutoff interval extends to higher and higher frequencies. The process reaches a physical limit when the frequency becomes so high that parasitic effects in the circuit are controlling and do not permit the prescribed cutoff

characteristic to be simulated with sufficient precision. For example, we are obviously in physical difficulties if the cutoff characteristic specifies a net gain around the loop at a frequency so high that the tubes themselves working into their own parasitic capacitances do not give a gain.

This limitation is studied most easily if the effects of the parasitic elements are lumped together by representing them in terms of the asymptotic characteristic of the loop as a whole at extremely high frequencies. An example is shown by Fig. 11. The structure is a

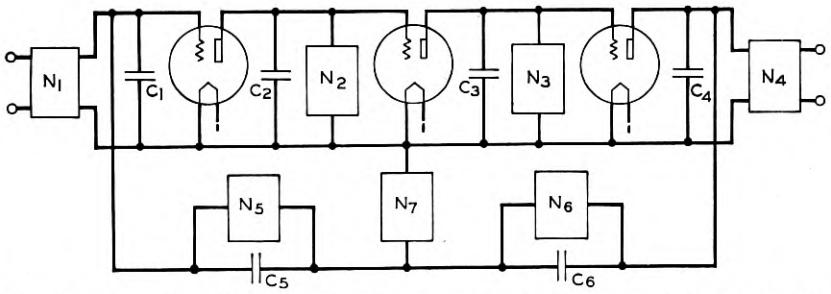

Fig. 11—Elements which determine the asymptotic loop transmission characteristic in a typical amplifier.

shunt feedback amplifier. The β circuit is represented by the T composed of networks N_5, N_6 and N_7. The input and output circuits are represented by N_1 and N_4 and the interstage impedances by N_2 and N_3. The C's are parasitic capacitances with the exception of C_5 and C_6, which may be regarded as design elements added deliberately to N_5 and N_6 to obtain an efficient high frequency transmission path from output to input. At sufficiently high frequencies the loop transmission will depend only upon these various capacitances, without regard to the N's. Thus, if the transconductances of the tubes are represented by G_1, G_2, and G_3 the asymptotic gains of the first two tubes are $G_1/\omega C_1$ and $G_2/\omega C_3$. The rest of the loop includes the third tube and the potentiometer formed by the capacitances C_1, C_4, C_5 and C_6. Its asymptotic transmission can be written as $G_3/\omega C$, where

$$C = C_1 + C_4 + \frac{C_1 C_4}{C_5 C_6}(C_5 + C_6).$$

Each of these terms diminishes at a rate of 6 db per octave. The complete asymptote is $G_1 G_2 G_3/\omega^3 C C_2 C_3$. It appears as a straight line with a slope of 18 db per octave when plotted on logarithmic frequency paper.

A similar analysis can evidently be made for any amplifier. In the particular circuit shown by Fig. 11 the slope of the asymptote, in units of 6 db per octave, is the same as the number of tubes in the circuit. The slope can evidently not be less than the number of tubes but it may be greater in some circuits. For example if C_5 and C_6 were omitted in Fig. 11 and N_5 and N_6 were regarded as degenerating into resistances the asymptote would have a slope of 24 db per octave and would lie below the present asymptote at any reasonably high frequency. In any event the asymptote will depend only upon the parasitic elements of the circuit and perhaps a few of the most significant design elements. It can thus be determined from a skeletonized version of the final structure. If waste of time in false starts is to be avoided such a determination should be made as early as possible, and certainly in advance of any detailed design.

The effect of the asymptote on the overall feedback characteristic is illustrated by Fig. 12. The curve *ABEF* is a reproduction of the ideal

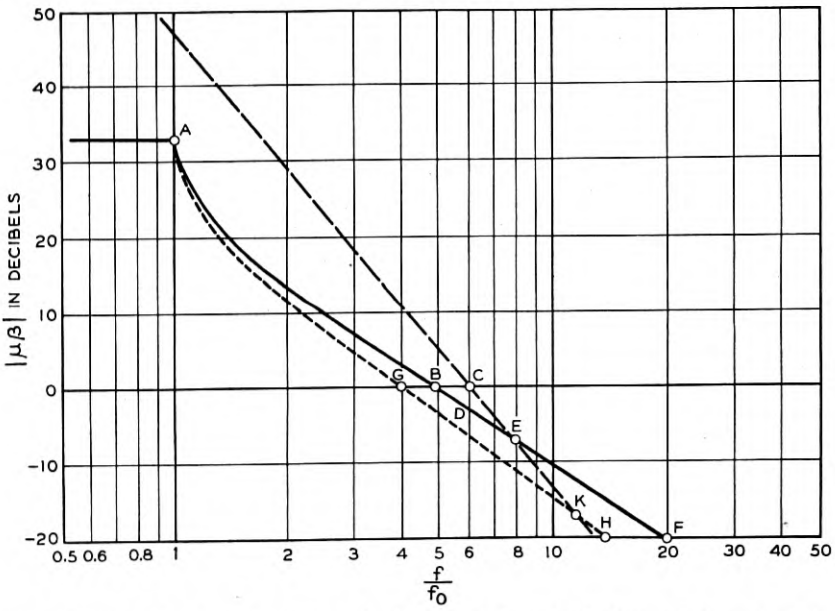

Fig. 12—Combination of asymptotic characteristic and ideal cutoff characteristic.

cutoff characteristic originally given by the solid lines in Fig. 10. It will be recalled that the curve was drawn for the choice $k = 5/3$, which corresponds to a phase margin of 30° and an almost constant slope, for the portion *DEF* of the characteristic, of about 10 db per octave. The

straight line *CEK* represents an asymptote of the type just described, with a slope of 18 db per octave. Since the asymptote may be assumed to represent the practical upper limit of gain in the high-frequency region, the effect of the parasitic elements can be obtained by replacing the theoretical cutoff by the broken line characteristic *ABDEK*. In an actual circuit the corner at *E* would, of course, be rounded off, but this is of negligible quantitative importance. Since *EF* and *EK* diverge by 8 db per octave the effect can be studied by adding curves of the type shown by Fig. 5 to the original cutoff characteristic.

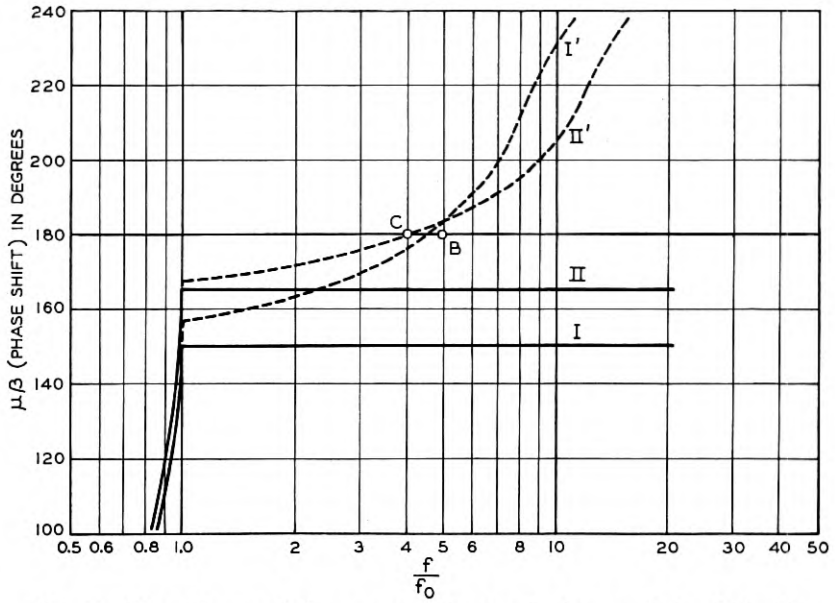

Fig. 13—Phase characteristics corresponding to gain characteristics of Fig. 12.

The phase shift in the ideal case is shown by Curve I of Fig. 13. The addition of the phase corresponding to the extra slope of 8 db per octave at high frequencies produces the total phase characteristic shown by Curve I′. At the point *B* where $|\mu\beta| = 1$, the additional phase shift amounts to 35 degrees. Since this is greater than the original phase margin of 30 degrees the amplifier is unstable when parasitic elements are considered. In the present instance stability can be regained by increasing the coefficient k to 1-5/6, which leads to the broken line characteristic *AGKH* in Fig. 12. This reduces the nominal phase margin to 15 degrees, but the frequency interval between *G* and *K* is so much greater than that between *B* and *E* that the added phase is reduced still more and is just less than 15° at the new

cross over point G. This is illustrated by II and II′ in Fig. 13. On the other hand, if the zero gain intercept of the asymptote CEK had occurred at a slightly lower frequency, no change in k alone would have been sufficient. It would have been necessary to reduce the amount of feedback in the transmitted range in order to secure stability.

The final characteristic in Fig. 13 reaches the limiting phase shift of 180° only at the crossover point. It is evident that a somewhat more efficient solution for the extreme case is obtained if the limiting 180° is approximated throughout the cutoff interval. This result is attained by the cutoff characteristic shown in Fig. 14. The characteristic con-

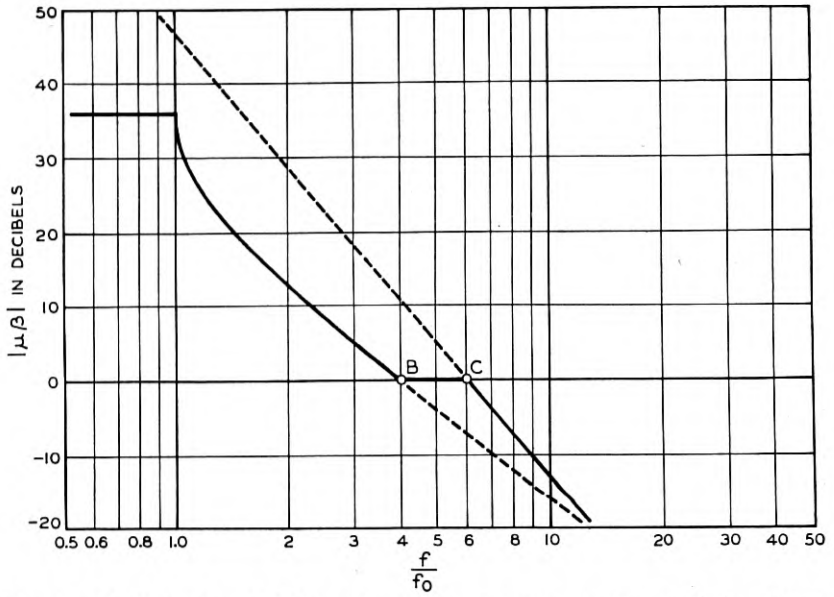

Fig. 14—Ideal cutoff modified to take account of asymptotic characteristic. Drawn for zero gain and phase margins.

sists of the original theoretical characteristic, drawn for $k = 2$, from the edge of the useful band to its intercept with the zero gain axis, the zero gain axis from this frequency to the intercept with the high-frequency asymptote, and the asymptote thereafter. It can be regarded as a combination of the ideal cutoff characteristic and two characteristics of the type shown by Fig. 5. One of the added characteristics starts at B and has a positive slope of 12 db per octave, since the ideal cutoff was drawn for the limiting value of k. The other starts at C and has the negative slope, − 18 db per octave, of the asymptote itself. As (3) shows, the added slopes correspond at lower frequencies to ap-

proximately linear phase characteristics of opposite sign. If the frequencies B and C at which the slopes begin are in the same ratio, 12 : 18, as the slopes themselves the contributions of the added slopes will substantially cancel each other and the net phase shift throughout the cutoff interval will be almost the same as that of the ideal curve alone. The exact phase characteristic is shown by Fig. 15. It dips

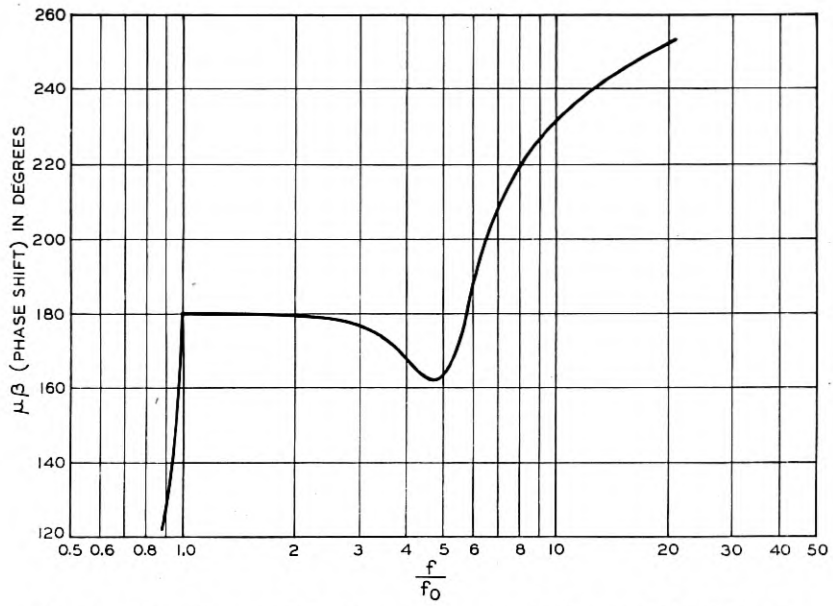

Fig. 15—Phase characteristic corresponding to gain characteristic of Fig. 14.

slightly below 180° at the point at which the characteristic reaches the zero gain axis, so that the circuit is in fact stable.

The same analysis can evidently be applied to asymptotes of any other slope. This makes it easy to compute the maximum feedback obtainable under any asymptotic conditions. If f_0 and f_a are respectively the edge of the useful band and the intercept (C in Figs. 12 and 14) of the asymptote with the zero gain axis, and n is the asymptotic slope, in units of 6 db per octave, the result appears as

$$A_m = 40 \log_{10} \frac{4f_a}{f_0}, \tag{7}$$

where A_m is the maximum feedback in db.[7]

[7] The formulæ for maximum feedback given here and in the later equation (8) are slightly conservative. It follows from the phase area law that more feedback should be obtained if the phase shift were exactly 180° below the crossover and rose

For the sake of generality it is convenient to extend this formula to include also situations in which there exists some further linear phase characteristic in addition to those already taken into account. In exceptional circuits, the final asymptotic characteristic may not be completely established by the time the curve reaches the zero gain axis and the additional phase characteristic may be used to represent the effect of subsequent changes in the asymptotic slope. Such a situation might occur in the circuit of Fig. 11, for example, if C_5 or C_6 were made extremely small. The additional term may also be used to represent departures from a lumped constant analysis in high-frequency amplifiers, as discussed earlier, If we specify the added phase characteristic, from whatever source, by means of the frequency f_d at which it would equal $2n/\pi$ radians, if extrapolated, the general formula corresponding to (7) becomes

$$A_m = 40 \log_{10} \frac{4}{nf_0} \frac{f_a f_d}{f_a + f_d} \cdot \tag{8}$$

It is interesting to notice that equations (7) and (8) take no explicit account of the final external gain of the amplifier. Naturally, if the external gain is too high the available μ circuit gain may not be sufficient to provide it and also the feedback which these formulæ promise. This, however, is an elementary question which requires no further discussion. In other circumstances, the external gain may enter the situation indirectly, by affecting the asymptotic characteristics of the β path, but in a well chosen β circuit this is usually a minor consideration. The external gain does, however, affect the parts of the circuit upon which reliance must be placed in controlling the overall loop characteristic. For example, if the external gain is high the μ circuit will ordinarily be sharply tuned and will drop off rapidly in gain beyond the useful band. The β circuit must therefore provide a decreasing loss to bring the overall cutoff rate within the required limit. Since the β circuit must have initially a high loss to correspond to the high final gain of the complete amplifier, this is possible. Conversely, if the gain of the amplifier is low the μ circuit will be relatively flexible and the β circuit relatively inflexible.

rapidly to its ultimate value thereafter. These possibilities can be exploited approximately by various slight changes in the slope of the cutoff characteristic in the neighborhood of the crossover region, or a theoretical solution can be obtained by introducing a prescribed phase shift of this type in the general formula (4). The theoretical solution gives a Nyquist path which, after dropping below the critical point with a phase shift slightly less than 180°, rises again with a phase shift slightly greater than 180° and continues for some time with a large amplitude and increasing phase before it finally approaches the origin. These possibilities are not considered seriously here because they lead to only a few db increase in feedback, at least for moderate n's, and the degree of design control which they envisage is scarcely feasible in a frequency region where, by definition, parasitic effects are almost controlling.

In setting up (7) and (8) it has been assumed that the amplifier will, if necessary, be built with zero margins against singing. Any surplus which the equations indicate over the actual feedback required can, of course, be used to provide a cutoff characteristic having definite phase and gain margins. For example, if we begin with a lower feedback in the useful band the derivative of the attenuation between this region and the crossover can be proportionately reduced, with a corresponding decrease in phase shift. We can also carry the flat portion of the characteristic below the zero gain axis, thus providing a gain margin when the phase characteristic crosses 180°. In reproportioning the characteristic to suit these conditions, use may be made of the approximate formula

$$A_m - A = (A_m + 17.4)y + \frac{n-2}{n}x + \frac{2}{n}xy, \tag{9}$$

where A_m is the maximum obtainable feedback (in db), A is the actual feedback, and x and y are the gain and phase margins in the notation of Fig. 9. Once the available margin has been divided between the x and

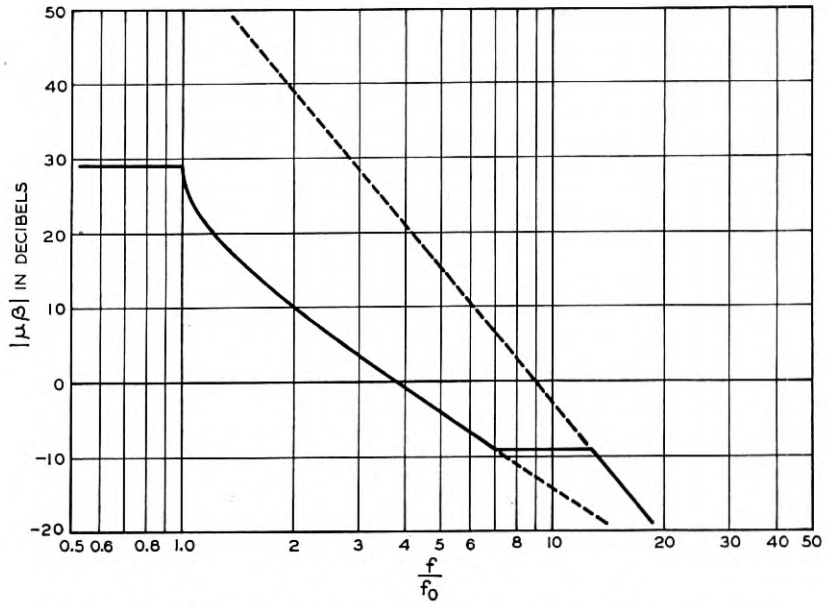

Fig. 16—Modified cutoff permitting 30° phase margin and 9 db gain margin.

y components by means of this formula the cutoff characteristic is, of course, readily drawn in. An example is furnished by Figs. 16 and 17,

where it is assumed that $A_m = 43$ db, $A = 29$ db, $x = 9$ db, $n = 3$ and $y = 1/6$. The Nyquist diagram for the structure is shown by Fig. 18. It evidently coincides almost exactly with the diagram postulated originally in Fig. 9.

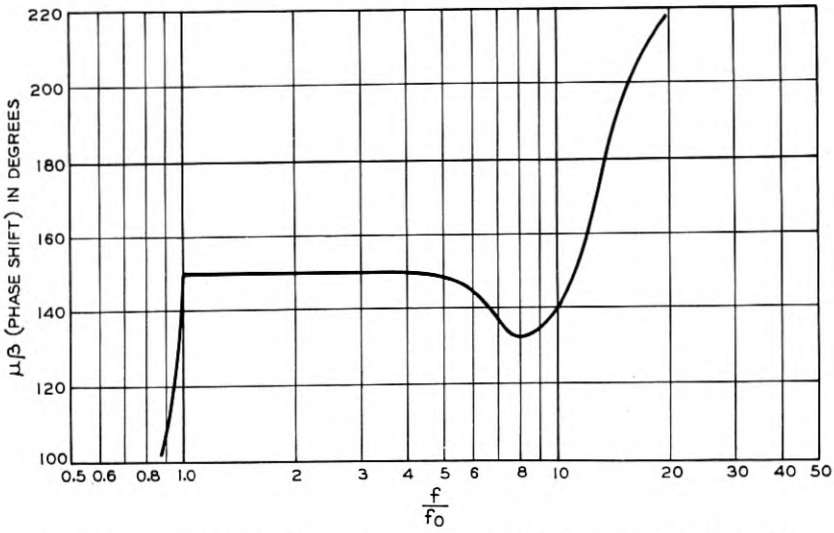

Fig. 17—Phase characteristic corresponding to gain characteristic of Fig. 16.

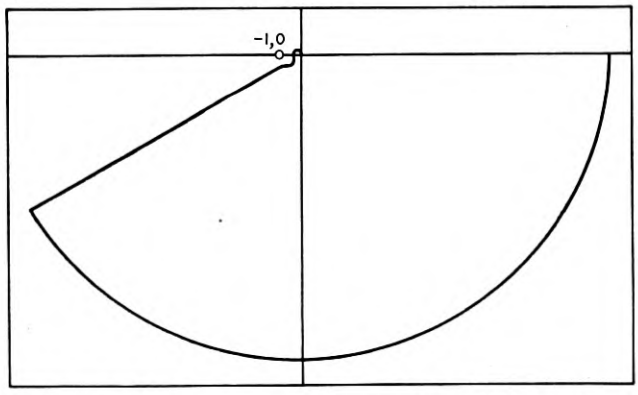

Fig. 18—Nyquist diagram corresponding to gain and phase characteristics of Figs. 16 and 17. As in Fig. 8 the diagram is rotated to place the critical point at $-1, 0$ rather than $+1, 0$.

With the characteristic of Fig. 16 at hand, we can return once more to the calculation of the total design range corresponding to any given feedback. From the useful band to the intersection of the cutoff

characteristic with the zero gain axis the calculation is the same as that made previously in connection with Fig. 10. From the zero gain intercept to the junction with the asymptote, where we can say that design control is finally relaxed, there is, however, an additional interval of nearly two octaves. Although Fig. 16 is fairly typical, the exact breadth of the additional interval will depend somewhat on circumstances. It is increased by an increase in the asymptotic slope and reduced by decreasing the gain margin.

Relative Importance of Tubes and Circuit in Limiting Feedback [8]

The discussion just finished leads to the general conclusion that the feedback which can be obtained in any given amplifier depends ultimately upon the high-frequency asymptote of the feedback loop. It is a matter of some importance, then, to determine what fixes the asymptote and how it can be improved. Evidently, the asymptote is finally restricted by the gains of the tubes alone. We can scarcely improve upon the result secured by connecting the output plate directly to the input grid. Within this limit, however, the actual asymptotic characteristic will depend upon the configuration and type of feedback employed, since a given distribution of parasitic elements may evidently affect one arrangement more than another. The salient circuit problem is therefore that of choosing a general configuration for the feedback circuit which will allow the maximum efficiency of transmission at high frequencies.

The relative importance of tube limitations and circuit limitations is most easily studied if we replace (7) by

$$A_m = 40 \log_{10} \frac{4f_t}{nf_0} - \frac{2A_t}{n}, \tag{10}$$

where f_t is the frequency at which the tubes themselves working into their own parasitic capacitances have zero gain [9] and A_t is the asymptotic loss of the complete feedback loop in db at $f = f_t$. The first term

[8] The material of this section was largely inspired by comments due to Messrs. G. H. Stevenson and J. M. West.

[9] I.e., $f_t = \frac{G_m}{2\pi C}$, where G_m and C are respectively the transconductance and capacitance of a representative tube. The ratio $\frac{G_m}{C}$ is the so-called "figure of merit" of the tube. The analysis assumes that the interstage network is a simple shunt impedance, so that the parasitic capacitance does correctly represent its asymptotic behavior. More complicated four-terminal interstage networks, such as transformer coupling circuits and the like, are generally inadmissible in a feedback amplifier because of the high asymptotic losses and consequent high-phase shifts which they introduce.

of (10) shows how the feedback depends upon the intrinsic band width of the available tubes. In low-power tubes especially designed for the purpose f_t may be 50 mc or more, but if f_0 is small the first term will be substantial even if tubes with much lower values of f_t are selected. The second term gives the loss in feedback which can be ascribed to the rest of the circuit. It is evidently not possible to provide input and output circuits and a β-path without making some contribution to the asymptotic loss, so that A_t cannot be zero. In an amplifier designed with particular attention to this question, however, it is frequently possible to assign A_t a comparatively low value, of the order of 20 to 30 db or less. Without such special attention, on the other hand, A_t is likely to be very much larger, with a consequent diminution in available feedback.

In addition to f_t and A_t, (10) includes the quantity n, which represents the final asymptotic slope in multiples of 6 db per octave. Since the tubes make no contribution to the asymptotic loss at $f = f_t$ we can vary n without affecting A_t by changing the number of tubes in the circuit. This makes it possible to compute the optimum number of tubes which should be used in any given situation in order to provide the maximum possible feedback. If A_t is small the first term of (10) will be the dominant one and it is evidently desirable to have a small number of stages. The limit may be taken as $n = 2$ since with only one stage the feedback is restricted by the available forward gain, which is not taken into account in this analysis. On the other hand since the second term varies more rapidly than the first with n, the optimum number of stages will increase as A_t is increased. It is given generally by

$$n = \frac{A_t}{8.68} \tag{11}$$

or in other words the optimum n is equal to the asymptotic loss at the tube crossover in nepers.

This relation is of particular interest for high-power circuits, such as radio transmitters, where circuit limitations are usually severe but the cost of additional tubes, at least in low-power stages, is relatively unimportant. As an extreme example, we may consider the problem of providing envelope feedback around a transmitter. With the relatively sharp tuning ordinarily used in the high-frequency circuits of a transmitter the asymptotic characteristics of the feedback path will be comparatively unfavorable. For illustrative purposes we may assume that $f_a = 40$ kc. and $n = 6$. In accordance with (7) this would provide a maximum available feedback over a 10 kc. voice band of 17 db. It

will also be assumed that the additional tubes for the low-power portions of the circuit have an f_t of 10 mc.[10] The corresponding A_t is 33 nepers [11] so that equation (11) would say that the feedback would be increased by the addition of as many as 27 tubes to the circuit. Naturally in such an extreme case this result can be looked upon only as a qualitative indication of the direction in which to proceed. If we add only 4 tubes, however, the available feedback becomes 46 db while if we add 10 tubes it reaches 60 db. It is to be observed that only a small part of the available gain of the added tubes is used in directly increasing the feedback. The remainder is consumed in compensating for the unfortunate phase shifts introduced by the rest of the circuit.

Amplifiers of Other Types

The amplifier considered thus far is of a rather special type. It has a useful band extending from zero up to some prescribed frequency f_0, constant feedback in the useful band, and it is absolutely stable. Departures from absolute stability are rather unusual in practical amplifiers and will not be considered here. It is apparent from the phase area relation that a conditionally stable amplifier may be expected to have a greater feedback for a cut-off interval of given breadth than a structure which is unconditionally stable, but a detailed discussion of the problem is beyond the scope of this paper.

Departures from the other assumptions are easily treated. For example, if a varying feedback in the useful band is desired, as it may be in occasional amplifiers, an appropriate cut-off characteristic can be constructed by returning to the general formula (4), performing the integrations graphically, if necessary. If the phase requirement in the cut-off region is left unchanged only the first integral need be modified. The most important question, for ordinary purposes, is that of determining how high the varying feedback can be, in comparison with a corresponding constant feedback characteristic, for any given asymptote. This can be answered by observing the form to which the first integral in (4) reduces when f_c is made very large. It is easily seen that the asymptotic conditions will remain the same provided the

[10] In tubes operating at a high-power level f_t may, of course, be quite low. It is evident, however, that only the tubes added to the circuit are significant in interpreting (11). The additional tubes may be inserted directly in the feedback path if they are made substantially linear in the voice range by subsidiary feedback of their own. This will not affect the essential result of the present analysis.

[11] It is, of course, not to be expected that the actual asymptotic slope will be constant from 40 kc. to 10 mc. Since only the region extending a few octaves above 40 kc. is of interest in the final design, however, the apparent A_t can be obtained by extrapolating the slope in this region.

feedback in the useful band satisfies a relation of the form

$$\int_0^{\pi/2} A d\phi = \text{constant}, \tag{12}$$

where $\phi = \sin^{-1} f/f_0$. Thus the area under the varying characteristic, when plotted against ϕ, should be the same as that under a corresponding constant characteristic having the same phase and gain margins and the same final asymptote. This is exemplified by Fig. 19, the

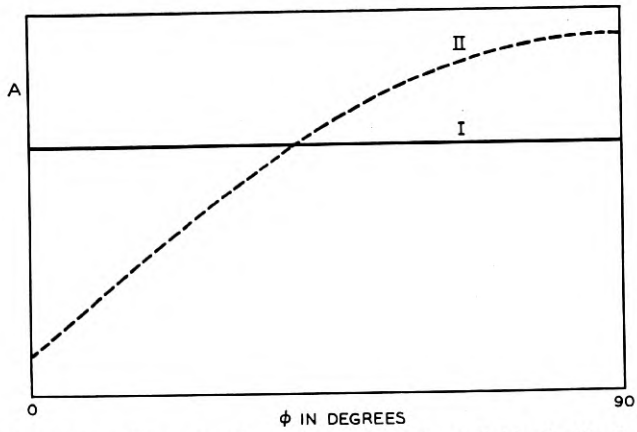

Fig. 19—Diagram to illustrate the computation of available feedback when the required feedback in the useful band is not the same at all frequencies.

varying characteristic being chosen for illustrative purposes as a straight line on an arithmetic frequency scale.

The most important question has to do with the assumption that the useful transmission band extends down to zero frequency. In most amplifiers, of course, this is not true. It is consequently necessary to provide a cut-off characteristic on the lower as well as the upper side of the band. The requisite characteristics are easily obtained from the ones which have been described by means of frequency transformations of a type familiar in filter theory. Thus if the cut-off characteristics studied thus far are regarded as being of the "low-pass" type the characteristics obtained from them by replacing f/f_0 by its reciprocal may be regarded as being of the "high-pass" type. If the band width of the amplifier is relatively broad it is usually simplest to treat the upper and lower cut-offs as independent characteristics of low-pass and high-pass types. In this event, the asymptote for the lower cut-off is furnished by such elements as blocking condensers and choke coils in the plate supply leads. The low-frequency asymptote is usually not so

serious a problem as the high-frequency asymptote since it can be placed as far from the band as we need by using large enough elements in the power supply circuits. The superposition of a low-frequency cutoff on the idealized loop gain and phase characteristics of a "low-pass" circuit is illustrated by the broken lines in Fig. 20.

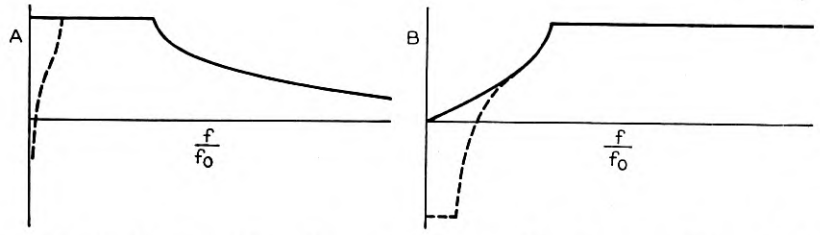

Fig. 20—Modification of loop characteristics to provide a lower cutoff in a broad-band amplifier.

If the band width is relatively narrow it is more efficient to use the transformation in filter theory which relates a low-pass to a symmetrical band-pass structure. The transformation is obtained by replacing f/f_0 in the low-pass case by $(f^2 - f_1f_2/f(f_2 - f_1))$, where f_1 and f_2 are the edges of the prescribed band. It substitutes resonant and anti-resonant circuits tuned to the center of the band for the coils and condensers in the low-pass circuit. In particular each parasitic inductance is tuned by the addition of a series condenser and each parasitic capacity is tuned by a shunt coil. The parameters of the transformation must, of course, be so chosen that the parasitic elements have the correct values for use in the new branches.

This leads to a simple but important result. If the inductance of a series resonant circuit is fixed, the interval represented by $f_b - f_a$ in Fig. 21, between the frequencies at which the absolute value of the reactance reaches some prescribed limit X_0, is always constant and equal to the frequency at which the untuned inductance would exhibit the reactance X_0, whatever the tuning frequency may be. The same relation holds for the capacity in an anti-resonant circuit. Thus the frequency range over which the branches containing parasitic elements exhibit comparable impedance variations is the same in the band-pass structure and in the prototype low-pass structure. But since the transformation does not affect the relative impedance levels of the various branches in the circuit, this result can be extended to the complete $\mu\beta$ characteristic. We can therefore conclude that *the feedback which is obtainable in an amplifier of given general configuration and with given parasitic elements and given margins depends only upon the breadth*

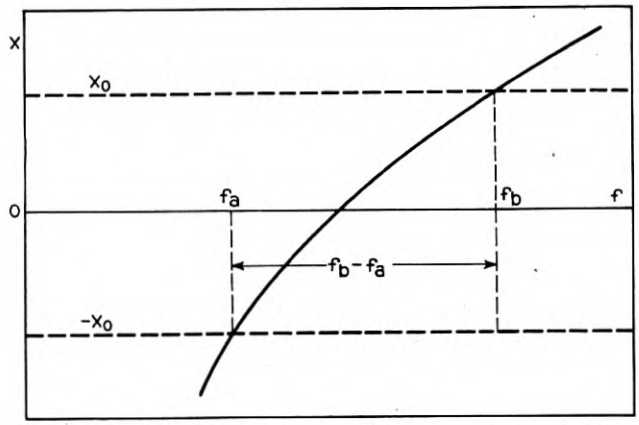

Fig. 21—Frequency interval between prescribed reactances of opposite sign in a resonant circuit with fixed inductance.

of the band in cycles and is independent of the location of the band in the frequency spectrum.

These relations are exemplified by the plots of a low-pass cutoff characteristic and the equivalent band-pass characteristic shown by Fig. 22. The equality of corresponding frequency intervals is indicated by the horizontal lines *A*, *B* and *C*.

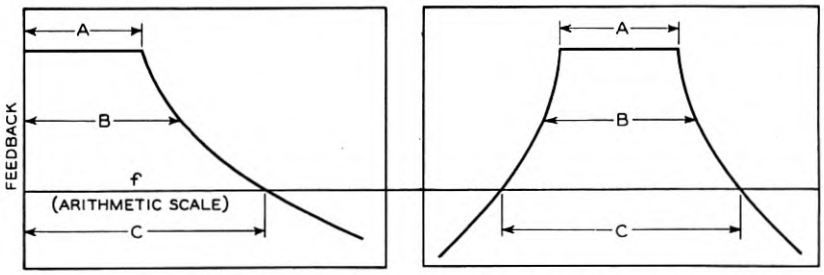

Fig. 22—Diagram to illustrate the conservation of band width in the low-pass to band-pass transformation with fixed parasitic elements. *A*, *B* and *C* represent typical corresponding intervals of equal breadth.

Example

An example showing the application of the method in an actual design is furnished by Fig. 23. The structure is a feedback amplifier intended to serve as a repeater in a 72-ohm coaxial line.[12] The useful frequency range extends from 60 to 2,000 kc. Coupling to the line is

[12] The author's personal contact with this amplifier was limited to the evolution of a paper schematic for the high frequency design. The other aspects of the problem are the work of Messrs. K. C. Black, J. M. West and C. H. Elmendorf.

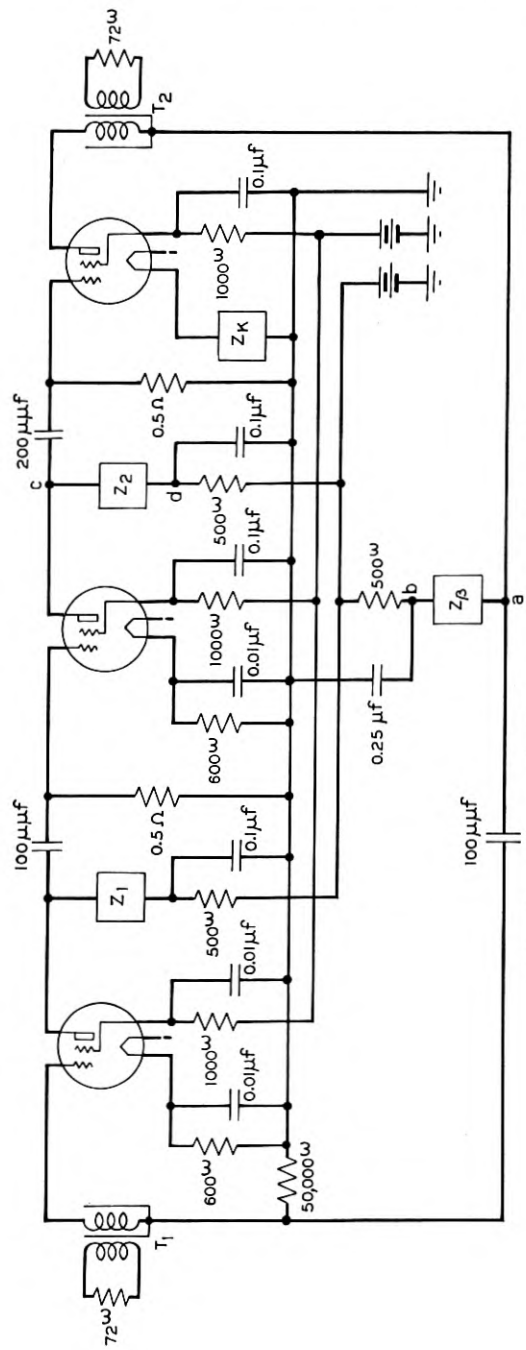

Fig. 23—General schematic of a 2 mc. feedback amplifier.

obtained through the shielded input and output transformers T_1 and T_2. The three stages in the μ circuit are represented in Fig. 23 as single tubes. Physically, however, each stage employes two tubes in parallel, the transconductances of the individual tubes being about 2000 micromhos. The principal feedback is obtained through the impedance Z_β. There is in addition a subsidiary local feedback on the power stage through the impedance Z_K. This is advantageous in producing a further reduction in the effects of modulation in this stage but it does not materially affect the feedback available around the principal loop.

The elements shown explicitly include resistance-capacitance filters in the power supply leads to the plates and screens, cathode resistances and by-pass condensers to provide grid bias potentials, and blocking condenser-grid-leak combinations for the several tubes. In addition to serving these functions, the various resistance-capacitance combinations are also used to provide the cutoff characteristic below the useful band. The low-frequency asymptote is established by the grid leak resistances and the associated coupling condensers and the approach of the feedback characteristic to the asymptote is controlled mainly by the cathode impedances and the resistance-capacitance filters in the power supply leads to the plates. The principal parts of the circuit entering into the $\mu\beta$ characteristic at high frequencies are the interstage impedances Z_1 and Z_2, the feedback impedance Z_β,[13] the cathode impedance Z_K, and the two transformers. The four network designs are shown in detail in Figs. 24, 25, 26, and 27.

The joint transconductance, 4000 micromhos, of two tubes in parallel operating into an average interstage capacity of 14 mmf, as indicated by Figs. 24 and 25, gives an f_t of about 50 mc. The parasitic capacities (chiefly transformer high side and ground capacities) in the other parts of the feedback loop provide a net loss, A_t, of about 18 db at this frequency. Since the asymptotic slope is 18 db per octave the intercept of the complete asymptote with the zero gain axis occurs about one octave lower, at slightly less than 25 mc. This is a relatively high intercept and may be attributed in part to the high gain of the vacuum tubes. The care used in minimizing parasitic capacities in the construction of the amplifier and the general circuit arrangement, including in particular the use of single shunt impedances for the coupling and feedback networks, are also helpful.

[13] The relative complexity of this network is explained by the fact that it actually serves as a regulator to compensate for the effects of changes in the line temperature. (See H. W. Bode, "Variable Equalizers," *Bell System Technical Journal*, April, 1938.) The present discussion assumes that the controlling element is at its normal setting. For this setting the network is approximately equal to a resistance in series with a small inductance. The fact that the amplifier must remain stable over a regulation range may serve to explain why the design includes such large stability margins.

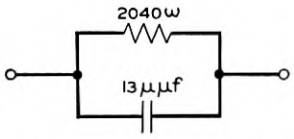

Fig. 24—First interstage for the amplifier of Fig. 23.

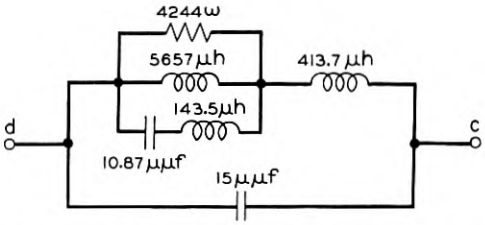

Fig. 25—Second interstage for the amplifier of Fig. 23.

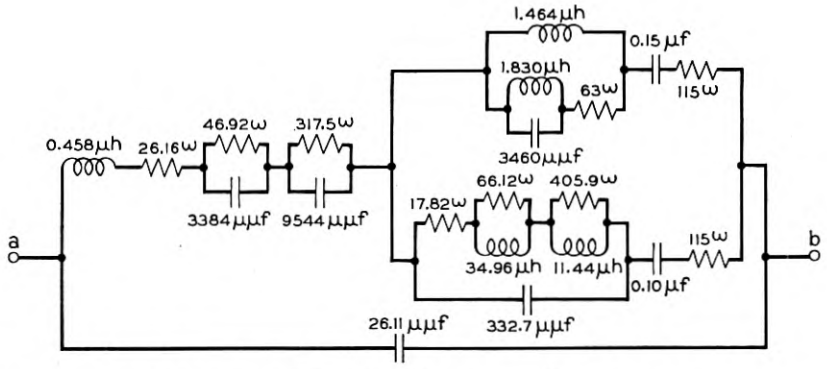

Fig. 26—β circuit impedance for the amplifier of Fig. 23.

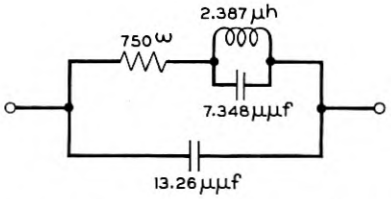

Fig. 27—Cathode impedance for the amplifier of Fig. 23.

In accordance with (7) the maximum available feedback A_m is 48 db. For design purposes, however, x and y in (9) were chosen as 15 db and 1/5 respectively. This reduces the actual feedback A to about 28 db. The theoretical cutoff characteristic corresponding to

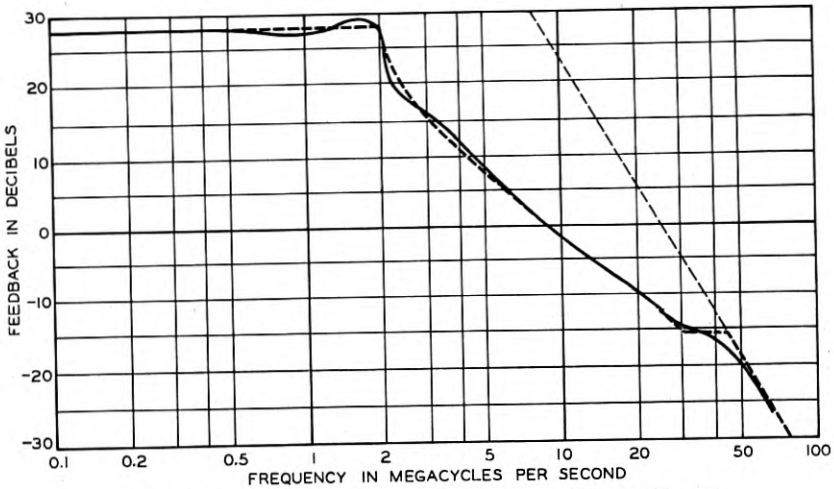

Fig. 28—Loop gain characteristic for the amplifier of Fig. 23.

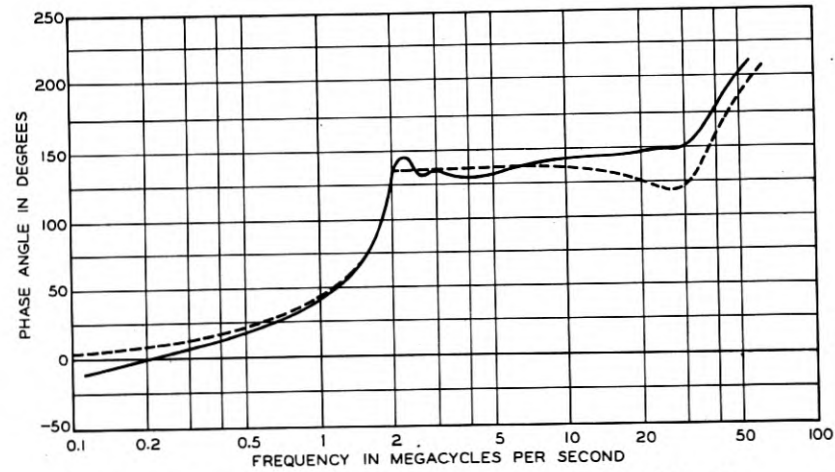

Fig. 29—Loop phase characteristic for the amplifier of Fig. 23.

these parameters is shown by the broken lines in Figs. 28 and 29, and the actual design characteristic by the solid lines. Since this is a structure in which the required forward gain is considerably less than the maximum available gain, the general course of the cutoff character-

istic is controlled, in accordance with the procedure outlined previously, by the elements in the μ circuit. The sharp slope just beyond the edge of the useful band is obtained from a transformer anti-resonance. The relatively flat portion of the characteristic near its intersection with the asymptote is due partly to an anti-resonance of the β circuit with its distributed capacitance and partly to an increase in the gain of the third tube because of the filter-like action of the elements of Z_K in cutting out the local feedback on the tube in this region.

The large margins in the design made it possible to secure a substantial increase in feedback without instability. For example, with a loss margin as great as 15 db the feedback can be increased by adjusting the screen and plate voltages to increase the tube gains. A higher feedback can also be obtained by adjusting the resistance in the first interstage. As this interstage was designed, an increase in the resistance results in an increased amplifier gain and a correspondingly increased feedback which follows a new theoretical characteristic with a somewhat reduced phase margin. The adjustment, in effect, produces a change in the value of the constant k in equation (6). With this adjustment the feedback can be increased to about 40 db before the amplifier sings.

Analysis of the Ionosphere *

By KARL K. DARROW

The ionosphere is a region in the very high atmosphere from which radio signals are reflected, a fact which is adequately explained by assuming that region to be populated with free electrons. In exploring the ionosphere, signals of a wide range of frequencies are successively sent upward, and the time elapsing before the return of the echo is measured. The delay of the echo mulitplied by $\frac{1}{2}c$ is called the virtual height of the ceiling for the signal. The data appear in the form of curves relating virtual height of ceiling to frequency of signal. These curves are peculiar in shape and vary remarkably with time of day, time of year and epoch of the solar cycle. By theory they can be translated into curves relating electron-density to true height above ground. The theory is approximative, but the results are accurate enough to be of value. The magnetic field of the earth affects the data remarkably, making it possible to test the theory and to evaluate the field-strength at great heights. The free electrons are supposed to be liberated from the air-molecules by ionizing agents, of which the chief but not the only one is ultra-violet light from the sun.

THE very title of this article embodies the assumption that in the upper reaches of the atmosphere there is a host of ions. By "upper reaches" here is meant, a region of the atmosphere so high that no man ever entered it, nor even a balloon with instruments. The ions therefore have never been observed by normal electrical means. They are postulated as the explanation of two things mainly: the echoing of radio signals from the sky, and that small portion of the earth's magnetic field which fluctuates with time.

The idea that these things require explanation, and the idea of the sort of postulate that is required to explain them, can both be followed back for many years. What was lacking in the early days was the notion of mobile electrified particles, that is to say, of "ions," in the air. That notion did not even exist, when in the eighties Balfour Stewart desired to imagine a conducting layer in the upper air for explaining magnetic fluctuations. It was only just being formed, when in 1902 Kennelly and Heaviside independently desired to imagine a conducting layer in the upper air for explaining why wireless signals can travel around the world. To speak of a conducting layer in the

* This paper, in abbreviated form, appears in the current issue of *Electrical Engineering*.

eighties, when air was regarded as an almost perfect insulator and nothing was known that could make it conductive, was certainly audacious. To speak of it in 1902 was still ingenious but no longer daring, for by then it could reasonably be expected that the researches on ions lately begun by Thomson and so many others would justify the notion.

Never was an expectation better founded. Within a few years those researches had made it sure that the upper atmosphere must be conductive, because of containing the raw material required for making ions and one at least among the agents capable of making them: to wit, atoms and molecules, and ultra-violet light from the sun capable of ionizing them. The problem then became: what distribution and what kinds of ions must be postulated for the upper atmosphere, to explain (for instance) the reflection of radio signals?

This problem could not even be attacked, without great forward strides in both the art of experimentation and the mathematical theory. These strides were rapidly made in the middle and late twenties. Had theory alone gone ahead, it would have been little more than a pretty exercise in mathematics. Had the art of experimentation progressed by itself, the experimenters would at least have found some interesting correlations of the data with such variables as time of day and epoch of the solar cycle and presence of magnetic storms; but the lack of theory would have been sorely felt. But theory and experiment advanced together, and the interplay between the two has seldom been so well exemplified.

The advance in the art of experiment lay not so much in the invention of new apparatus (though this has not been wanting) as in turning away from the practical problem of sending signals to great distances, and instead designing the experiments for the purpose, first of proving the ionosphere and then of "sounding" it. Three methods were invented for this purpose, all based upon the fact that wireless waves when sent into the sky come bouncing back from it. Two of these will be scarcely more than mentioned in the pages to follow, since an already great and ever-increasing proportion of the data is obtained by the third. In this third a sharply-delimited signal or pulse or wave-group is sent up, and a short time (a few milliseconds) later it is detected coming back, like an echo from a cliff: the delay of the echo is measured. This is done for many signals, and the delay is plotted against the mean frequency of the wireless waves composing the signals; and curves so plotted constitute the ultimate data. Usually the signal is sent vertically upward, the echo comes vertically downward; and there is the quaint situation, that wireless telegraphy

is chiefly famed for bridging great distances over the earth, but its foundations are best studied with sender and receiver side by side.

Electromagnetic signals thus find a mirror or a ceiling overhead; and the theory interprets this mirror as consisting of the ions, and especially the free electrons, diffused in the upper air. This perhaps seems singular, in view of the tenuity of the air and the lightness of the individual electrons. It might have seemed better, at least in the days of the Greeks, to propose that the dome of the sky is a hard metallic mirror—of well-polished silver, for instance. Well, in effect that is what *is* proposed. A mirror of silver reflects not by virtue of its hardness, but because of electrons diffused like a gas through the pores of the metal. A metal is a container for an electron-gas, and in the upper air there is an electron-gas without a container; and both of them reflect.

The theory is strictly classical, in the sense of the word prevailing in physics. No relativity, no quantum theory, no suggested revision of the concepts of space and time, afflict the student thereof. It is the working-out of the basic principle of Maxwell and Lorentz, that the passage of electromagnetic waves through a medium is controlled by the electric current which the waves themselves evoke in the medium. Under the influence of the electric field in the waves, the ions swing in sympathetic vibration, and form a part of that current. They thus react upon the waves, alter the speed thereof, and bring about the reflection. The motion of the ions is simple-harmonic, so that the mathematics of the theory is simple and familiar—so long, at least, as no account is taken of any forces acting on the ions other than that due to the field of the waves themselves. Here is the explanation of the echoing of radio signals, and hence follows the procedure for translating the data of echoes into statements about the distribution of the ions in the atmosphere. It is not difficult to describe or explain, and will be carried through in this article.

From this point the theory ramifies in two directions. Two things modify the sympathetic vibrations of the ions: the collisions between the ions and the neutral molecules of the air, and the earth's magnetic field. By their influence on the vibrations, they modify the speed of the waves, and therefore the conditions of the echoing. The theory extended in either direction continues to be easy in one sense, for neither the physical concepts nor the mathematical operations are unfamiliar; but becomes very hard in another, for the algebraic expressions are often of fearful complexity, impossible to remember and hard even to keep straight when written out. When it is extended in both ways at once the expressions become so intricate, that nearly

every investigator when taking account of either influence simply ignores the other. On the whole, the mathematical developments have far outrun the data. Yet there are important connections between experiment and theory, including for instance the proof that the ions which principally reflect the signals are free electrons.

After the theory come what I will call, for contrast, the speculations. The analysis of the ionosphere being made and accepted, a host of questions arise. Must we assume additional agents of ionization, other than the ultra-violet light of the sun? The answer to this question being certainly "yes," one must inquire how to distinguish that part of the ionization which is due to sunlight from the rest, and what are the causes of the rest. Why the distinctive distribution-in-height of the ions, amounting to what is called "the stratification of the ionosphere"? What assumptions must we make about the composition of the atmosphere in its dependence on height? or (as the question is more commonly put) what information can we derive about the composition of the atmosphere? How far can we go in interpreting the fluctuations of terrestrial magnetism, and (as later will be apparent) in mapping out the earth's magnetic field? The possible questions even rise to the realm of astronomy, and the suggested answers form a part of the theory of the sun as a potent source of radiations of all kinds, luminous and electrical and material. The implications of the ionosphere seem to be almost limitless, but a severe limit will nevertheless be set by space and time upon this article.

Methods of Experiment, and a Simplified Picture of the Ionosphere Adduced for Illustrating Them

The ionosphere is a canopy of ions overarching the earth, and in Fig. 1 it is represented by a model, very simplified indeed and yet instructive. Here it is shown as consisting of two "layers" marked E and F, with an ion-density which is uniform in each, and greater in F than in E. It is time to become familiar with the symbol N used for number of ions per unit volume: this picture shows N having the constant values N_E and N_F ($> N_E$) in E-layer and F-layer respectively, and the value zero between.

The lines which are broken at the layer-edges are paths of wireless signals or waves sent out from the source at S—sent out obliquely, for transmission over long distances. There is a path reflected from E, a path reflected from F and a path which penetrates both. These correspond to relatively low, medium, and high frequencies respectively: as examples I will give the values 1, 10 and 100 mc. (megacycles,

i.e. millions of cycles per second) corresponding to wave-lengths of 300, 30 and 3 metres. Here already the reader meets the fact that the height at which such a signal is reflected, or the question whether it shall be reflected at all, depends on the frequency of the waves and the density of the ions. For every frequency there is what I shall call a "mirror-density": signals are reflected as soon as they reach the lowest level in the ionosphere where that mirror-density is attained. The higher the frequency, the higher the mirror-density. The formula

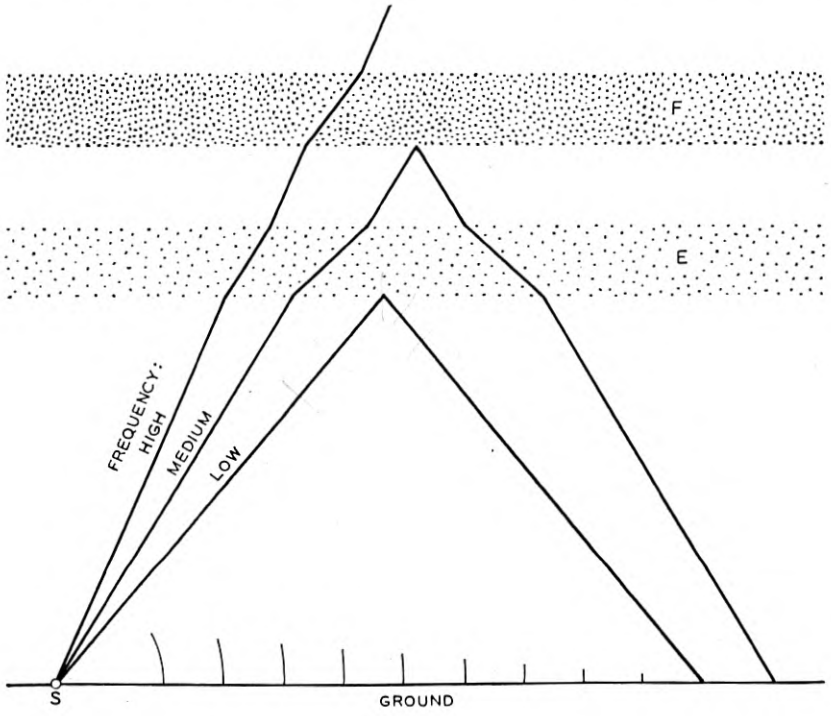

Fig. 1

will soon be derived and shown; but for the moment, let us inquire how the altitude of either layer can be measured, in terms of the simple model of Fig. 1.

One way of measuring the altitude is very obvious. Suppose the observer to go a known distance away from the aerial, and measure there the angle which the reflected wave or "sky wave" makes with the horizontal as it comes down to him. If this can be done, then clearly he can get the altitude by the simplest trigonometry—the altitude of the *E*-layer or the *F*-layer, according to the frequency

which he uses. It *can* be done, and it was done by Appleton and Barnett in 1925. What they measured was the angle between the directions of propagation of the sky wave and the "ground wave," of which the wave-fronts are shown in Fig. 1 creeping along the ground with a rapid attenuation. This ground wave, by the way, is the only one by which radio transmission could be effected but for the ionosphere; and it is seldom detectable beyond a few hundred miles.

Another scheme is much more complicated, and owing to its supersession I may be excused for giving only the merest outline of it. It is a clever way of putting to useful service the very great inconvenience known as "fading." This term refers to a train of signals which dies out and revives and keeps on fluctuating over and over again, in a most irregular fashion. This sort of thing occurs in the region where the sky wave and the ground wave both arrive and overlap one another, and it has been traced to what in optics is called the "interference" of the two. If conditions were absolutely stable, then in taking a walk in the region of overlapping one would pass through several maxima and minima of intensity. Since conditions are never absolutely stable, the observer need not take the walk; while he stands at any fixed point, the maxima and the minima float past him while the ionosphere wavers in the sky and this is "fading." But imagine the conditions relatively stable for a time, and the observer standing still; and suppose that the engineer at the sending aerial changes the wave-length by a small and known amount—then, several maxima and minima will float past the observer, and by counting them he can (though this is not at all obvious!) get a datum which enables him to figure out the altitude of the reflecting layer in the sky. This method also was invented by Appleton, and can be found explained in the literature under the name "wave-length-change method."

The third of the methods has crowded out the others, and henceforth will figure alone in these pages. It is the "echo-method," still sometimes called by the clumsy name of "group-retardation method." Anticipated by Swann, it was realized by Breit and Tuve at the Carnegie Institution of Washington.

What is sent up to the sky is here a short sharp signal; if it could be heard, it would be called a click. What comes back is the echo of the signal. Passing over the receiving device, it produces a short sharp kick on an oscillograph-record. Some of these are shown in Fig. 2. The kicks marked E_1 and F_1 are due to signals echoed from the layers E and F respectively. Those marked $F_2 \cdots F_5$ are due to

"multiple echoes"; the signal has traveled two to five times the entire journey from ground to ionosphere to ground again, the surface of the earth being itself a good reflector. Those marked *G* are due to the signal spreading along the ground itself. If the sender and the receiver are practically side by side, as usually is the case, the kicks *G* occur at the instants of departure of the signals. The record is moving laterally with the speed intimated by the wavy line beneath, and accordingly the distance along it from a *G*-kick to the following echo-kick is a measure of the "delay of the echo."

The delay of the echo is an indication of the altitude of the mirror where it was reflected—the layer *E* or *F*, as the case may be. Signals of relatively low frequency being reflected from *E* while those of medium frequency are echoed at *F*, one adjusts the frequency according

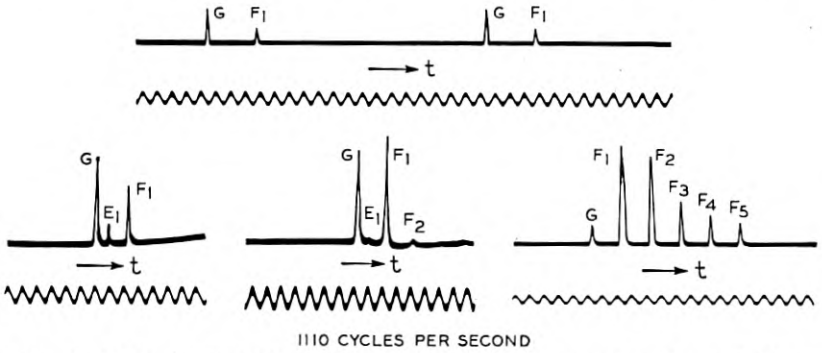

Fig. 2—Echoes. *G*, original signal; E_1, F_1, echoes returning after a single reflection from *E* and *F* respectively; $F_2 \cdots F_5$, echoes which have suffered two to five reflections at *F*-layer. (Appleton and Builder.)

to the layer which one wishes to locate. If there should be not two but several layers of the ionosphere, each having a greater *N*-value than the one beneath it, one would locate them all with appropriate frequencies. If there is a continuously-rising distribution of *N* with height in the ionosphere, one may plumb it by varying the frequency continuously. Now we are at the principle of the echo-method; but before it is used, there are many details to clear up.

First as to the "signals," a term which (it must have been noticed) replaced the term "waves" in the foregoing paragraphs. These signals are wave-trains indeed, but not the long continuous uniform trains tacitly assumed in the description of the other two methods. Those methods are adapted to trains of indefinite length, but not the echo-method, which for an obvious reason requires wave-trains of limited length—and the more limited, the better.

Now, a limited wave-train is equivalent to an infinity of infinitely long wave-trains, with an infinite variety of frequencies. The amplitudes and the frequencies of these constituent waves are chosen so that the waves reinforce one another over the length of the signal, counteract one another for all time before and after the signal. To choose them thus is always mathematically feasible, whatever shape of signal be prescribed. Whether these constituent waves should be regarded as "physically real" is a question that was discussed long before the days of quantum theory and other modern puzzles. Anyhow, by taking them as such, one arrives at verifiable statements about the signals, and this is all that matters.

Let us now conceive the signal as a chopped-off segment of a sine-wave-train of frequency f_0; and let us compare its travel with the travel of a limitless wave-train of frequency f put equal to f_0. One feels that the signal ought to follow the same path through the ionosphere as would the limitless train, and ought to move along that path with the same speed as would the wave-crests of the limitless train. This is true if, but only if, the speed u of the wave-crests in unlimited trains is independent of f. But when wireless waves are traveling through the ionosphere, u varies very much with f, according to a law which will later be worked out; and this makes a remarkable difference.

The difference as to path is not serious. The infinite wave-trains which form the signal are most intense at frequencies very close to f_0, and this is sufficient to make the signal follow nearly (though not without some deviation and distortion) the path which the infinite train of frequency f_0 would follow by itself. We may therefore regard the broken lines of Fig. 1 as the paths of signals or of waves, indifferently.

The difference as to speed *is* serious. It is not reduced by the preponderance of component wave-trains very close to f_0, and it does not tend to vanish as this preponderance is increased by lengthening the signal. It remains serious in the artificially-simplified case of just two component wave-trains of small frequency-difference Δf, where the signals become the "beats" well known in acoustics and in radio. In this case the beat-speed or signal-speed approaches a limit as Δf approaches zero. This limit, the "group-speed" denoted by v, is always used for the signal-speed, though for actual signals it is but an approximation, and the signals themselves become distorted as they travel. Contrasted with v is the "wave-speed" or the "phase-speed" of the wave-crests of the unlimited wave-trains, already denoted by u.

Since it is entirely the signal-speed which determines the delay of the echo, the wave-speed may seem a pointless side-issue. This quantity u is, however, essential in the theory from which are derived first the conditions of echoing and then the dependence of v upon f_0. Postponing the topics of signal-speed and echo-delay in order to return to them later with better preparation, we now take up the theory.

Theory of Wave-Speed, Total Reflection and Group-Speed in the Ionosphere

It will now be proved from Maxwell's theory, combined with the concept of mobile ions, that total reflection of wireless waves must occur in the ionosphere at the level where the ion-density attains a certain value depending on the wave-frequency.

The famous equations of Maxwell melt together into a wave-equation. The waves which it describes consist of an oscillating electric field which I will denote by $E_0 \sin nt$, and an oscillating magnetic field which we are permitted to ignore. When of high enough frequency these are the waves of light, as Maxwell knew; when of the frequency-range with which we are now concerned they are the waves of radio, as Maxwell was never to know because of his premature death. In the wave-equation there figures of course the wave-speed u. Here then is a paraphrase of the great idea of Maxwell: *the square of the wave-speed varies inversely as the current-density provoked by unit amplitude* [1] *of the oscillating field.*

Now we see at once that in the ionosphere the wave-speed must be affected by the presence of the free electrons, since they are set into oscillation by the waves and therefore make a contribution to the current-density.

At this point those who were educated in the electronic era (an ever-increasing fraction of the population) are in some danger of falling into a serious error. One may in fact assume that the electrons form the whole of the current, and deduce that in vacuo the oscillating field provokes no current at all, and the wave-speed must therefore be infinite—an absurd conclusion! Maxwell was wiser. He understood, and made it a part of his theory, that wherever there is an electric field which is changing in time the rate-of-change of that field is equivalent to a current. This he called the "displacement-current," and for the case of vacuum he said that the displacement-current-density is precisely equal to the rate-of-change of the field, multiplied by $1/4\pi$.

[1] I introduce the words "unit amplitude" to shield the reader from drawing the false inference that wave-speed depends upon wave-amplitude.

In vacuo, therefore, there is a current-density $(1/4\pi)$ times the rate-of-change of $E_0 \sin nt$, and it has an amplitude of $(nE_0/4\pi)$ and is 90° ahead of the field in phase. To this current-density corresponds the speed of light in vacuo, the well-known constant c. The speed of light in the non-ionized lower regions of the atmosphere differs so little from c that we need never bother with the difference, which henceforth will be ignored.

When the waves pass out of ordinary air into the ionosphere, there is still the displacement-current but now in addition there is the current borne by moving electrons. Here is a second pitfall. It may seem obvious that the electron-current must add on to the displacement-current, creating a total current-density greater than that in vacuo and therefore lowering the wave-speed. Not so at all! The point is, that when the electrons are truly free, the field sets them into oscillation in such a curious way that when they become adjusted, they are oscillating with their velocities 90° behind the field in phase. Their contribution to the current, being proportional to their velocity, is also 90° behind the field, and hence in perfect opposition of phase to the displacement-current.

Therefore the electron-current density—call it I_e—is to be *subtracted* from the displacement current-density! Accordingly I write,

$$\frac{u^2}{c^2} = \frac{n(1/4\pi)E_0}{n(1/4\pi)E_0 - I_e}. \tag{1}$$

The reader may suppose that the factor $\cos nt$, common to both currents, has been divided out.[2] The quantity I_e is clearly proportional to E_0 and also to our familiar N the density of electrons, and in fact the reader can undoubtedly work out with ease that it is equal to NE_0e^2/mn. Here e and m stand for the charge and mass of the ion, as is customary. Therefore we find:

$$\frac{u^2}{c^2} = \frac{1}{1 - 4\pi Ne^2/mn^2} = \frac{1}{1 - Ne^2/\pi mf^2}. \tag{2}$$

The wave-speed is greater in the ionosphere than it is in vacuo or ordinary air. I now recall from the most elementary optics the principle that when two media adjoin in which light has different wave-speeds, and light passes through their common boundary into the medium where its speed is greater, it is refracted away from the normal to the boundary. Accepting for the moment the over-simplified model of the

[2] Actually Maxwell's theorem does refer to the amplitudes of the currents—but if the currents are not exactly 0° or 180° apart in phase, the amplitude of one must be taken as a complex quantity.

ionosphere in Fig. 1, and considering the lower frequencies, we have the non-ionized lower atmosphere and the E-layer for these media. The paths of the waves are drawn accordingly. Now I further recall that total reflection occurs for all values of the angle of incidence i greater than that given by the equation:

$$\sin i = c^2/u^2. \tag{3}$$

Thus we see that for any frequency whatever, total reflection must occur when the waves impinge with sufficient obliqueness upon the ionosphere; but (so long as c/u does not sink to zero) total reflection will not occur if the waves rise vertically, or in a direction sufficiently near to the vertical.

The waves thus penetrate or are reflected back from the ionosphere, according as their angle of incidence thereon is less or greater than a certain critical value. Here is the explanation of what is called "skip-distance": the sky-wave is perceived beyond a certain distance from the source, but not within that certain distance.[3]

But all this seems to have nothing to do with the usual conditions of experiment, in which, as I intimated, the signals are sent up vertically! It is indeed a fact that in optics, no case is known in which total reflection occurs at vertical incidence. Yet equations (2) and (3) predict that if ever c^2/u^2 should vanish, total reflection would extend even to vertical incidence. Now there is nothing mathematically impossible or physically unplausible about the condition for the vanishment of c^2/u^2, which is simply that f should be equal to f_c given thus:

$$f_c^2 = Ne^2/\pi m \tag{4}$$

or alternatively that N should be equal to N_c given thus:

$$N_c = \pi m f^2/e^2. \tag{5}$$

Here we have the basic formula of the analysis of the ionosphere; for it is assumed that vertically-rising waves or signals of any frequency f climb until they reach the lowest level at which N is equal to N_c, and there they find their mirror or their ceiling, and are converted into echoes which return. Equation (5) is the formula for the "mirror-density" for signals of frequency c, to which I above referred.

It sounds all right to say that c^2/u^2 is zero when $N = N_c$, and negative when $N > N_c$; but it is disconcerting to notice that this

[3] Notice incidentally that owing to the curvature of the earth and its overhanging ionosphere, the angle of incidence can never rise to 90°; it follows that waves of frequency beyond a certain value (ordinarily around 30 mc.) never suffer total reflection.

amounts to saying that the phase-speed is infinite when $N = N_c$, imaginary when $N > N_c$. However, the concept of phase-speed is of such a quality of abstractness, that even these statements imply nothing absurd in the physical situation. The signal-speed itself remains safely finite and real.

The signal-speed is strictly indefinite, since the signal distorts itself as it proceeds. However, the practice is to identify it with the group-speed v, which, as I intimated (page 462), is the speed of the beats formed by two superposed wave-trains differing infinitesimally in wave-length, each such beat being a very special type of signal. The formula is,

$$v = u - \lambda(du/d\lambda) = u \Big/ \left(1 - \frac{n}{u}\frac{du}{dn}\right). \tag{6}$$

It is difficult to visualize or derive without a diagram,[4] but the derivation may be summarized as follows. Imagine two superposed wave-trains of phase-speeds u and $u + du$, wave-lengths λ and $\lambda + d\lambda$; consider two consecutive wave-crests A, A' of one and two consecutive wave-crests B, B' of the other; transpose temporarily to a frame of reference in which the former wave-train is stationary. At a certain place and time A and A' will coincide, and the maximum of one of the beats will be right there. Let the time $d\lambda/du$ elapse; when it has elapsed, the crests B and B' will be coinciding and the maximum of the beat will have moved on by one entire wave-length. The beat therefore travels with speed $\lambda du/d\lambda$ in the temporary and with speed $u - \lambda(du/d\lambda)$ in the original frame of reference (the minus sign is evident when the reasoning is gone through in detail).

Combining (6) with (2) one finds:

$$v = c^2/u; \tag{7}$$

the greater the phase-speed, the slower the signal! Relativists will be pleased to observe that according to this formula, the signal never attains any speed greater than c; students of quantum mechanics may be misled by its superficial resemblance to a formula relating phase-speed to group-speed for de Broglie waves, with which it has nothing to do. Students of the ionosphere should remember its approximative character. Almost all that needs to be known for the purposes of this article is, that as a signal climbs into the ionosphere it goes more and more slowly, the nearer N approaches to that value N_c where the signal finds its ceiling.

[4] Cf. this journal, 9, 173 (1930), or my *Introduction to Contemporary Physics*, 2nd edition, p. 147.

CHARACTERISTIC CURVES OF THE IONOSPHERE: THE (h', f) CURVES

Now that we have the concept of a signal ascending until it reaches the ceiling where $N = N_c$, we will consider first the to-be-expected relation between true height of ceiling and frequency of signal, then the relation between delay of echo and frequency of signal. By following this order we pass from the unobservable to the observed, which is the reverse of the customary way, but nevertheless has its advantages. Let z stand for height over ground when used as independent variable, with N depending on it; h for the height of the mirror or ceiling for signals, when expressed as a function of the signal-frequency f.

Take first the oversimplified model of the ionosphere appearing in Fig. 1: N having the values N_E over one range of heights and N_F ($> N_E$) over another range at a higher elevation, and the value zero elsewhere. It is evident that the (h, f) curve for such an ionosphere would consist of two horizontal lines or "branches," extending respectively from abscissa 0 to abscissa $f_E = \sqrt{N_E e^2/\pi m}$ and from abscissa f_E to abscissa $f_F = \sqrt{N_F e^2/\pi m}$ respectively. The latter would lie higher than the former; there would be a jump or gap between the branches. The names "E-branch" and "F-branch" for these last are obvious, and so is the usage "penetration-frequency of the E (or F) layer" for f_E or f_F; "critical frequency" is also used.

Next we approach closer to the truth by supposing that N rises continuously with increase of z across the E-layer and also across the F-layer, N_E and N_F now representing the highest N-values found in the respective layers. The two branches of the (h, f) curve would then be no longer horizontal, but slanting or probably curving upwards toward the right.

In the foregoing paragraph it was tacitly assumed that N still vanishes between the layers; but now let us approach still closer to the truth by postulating the sort of dependence of N on z shown in Fig. 3*A*. Here N drops with further increase of height after the "crown" of the E-layer is reached, but it does not fall to zero. It might, however, just as well fall to zero so far as reflections are concerned, for the signals which could be reflected from these regions never reach them. Regarding the curve of Fig. 3*A* as a sequence of hills and valleys, we see that the valleys contribute no echoes. The E-branch of the (h, f) curve refers to the left-hand side of the first hill; the F-branch refers to the left-hand side of the second hill, and not even to all of that, but only to the portion which rises above the first hill. Thus Fig. 3*B*, with its upturning branches and its gap, represents the (h, f) curve for the ionosphere of Fig. 3*A*, without in

the least depending on the dashed parts of the $N(z)$ curve of Fig. 3*A* or indicating anything whatever about those parts except that they do not rise above the ordinate N_E.

Now if the signal and the echo traveled fro and to with the speed c, the delay T of the echo multiplied by $\frac{1}{2}c$ would be the height of the ceiling. This, however, is not the case, since the signal-speed depends on N. We must therefore denote the product $\frac{1}{2}cT$ by another symbol h', and make an inquiry into the probable dependence of h' on f, taking into account our vague knowledge as to the dependence of signal-speed on N.

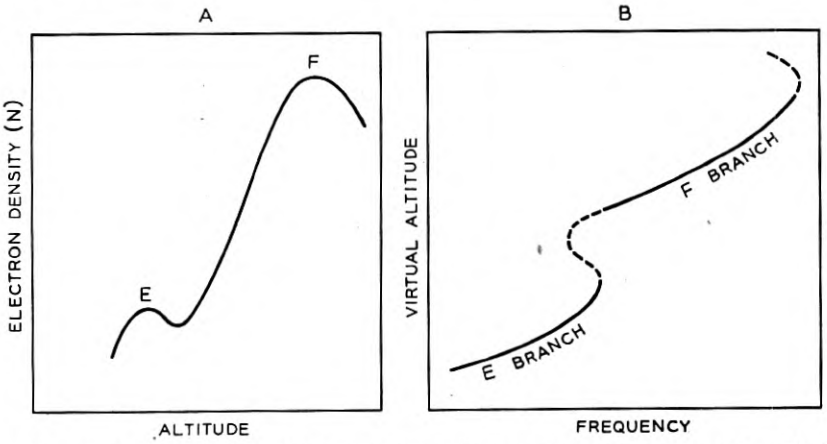

Fig. 3—*A*. The "curve of inference": conjectural dependence of number N of electrons per unit volume on true altitude h. *B*. The "curve of data": dependence of virtual altitude h' of ceiling (one-half the delay of the echo, multiplied by c) on frequency f.

It is easily seen that h' must be greater (or at least no less) than h, and that the excess of h' over h must be larger, the farther the signal travels through regions where N is almost but not quite equal to N_c. The (h', f) curve must therefore lie above the (h, f) curve, and farthest above it in the immediate neighborhood of the gap on both sides. There will still be an E-branch and an F-branch, but the upturns toward the right-hand ends of these branches will be exaggerated, and an upturn running to the left will be introduced into the left-hand end of the F-branch. It is conceivable that these upturns may become so large, that the (h', f) curve will appear to show a peak where the (h, f) curve would show a gap.

With the remark that h' is known as "virtual altitude," "virtual height," "equivalent height," or "effective height," I turn now to examples of the characteristic (h', f) curves of the ionosphere.

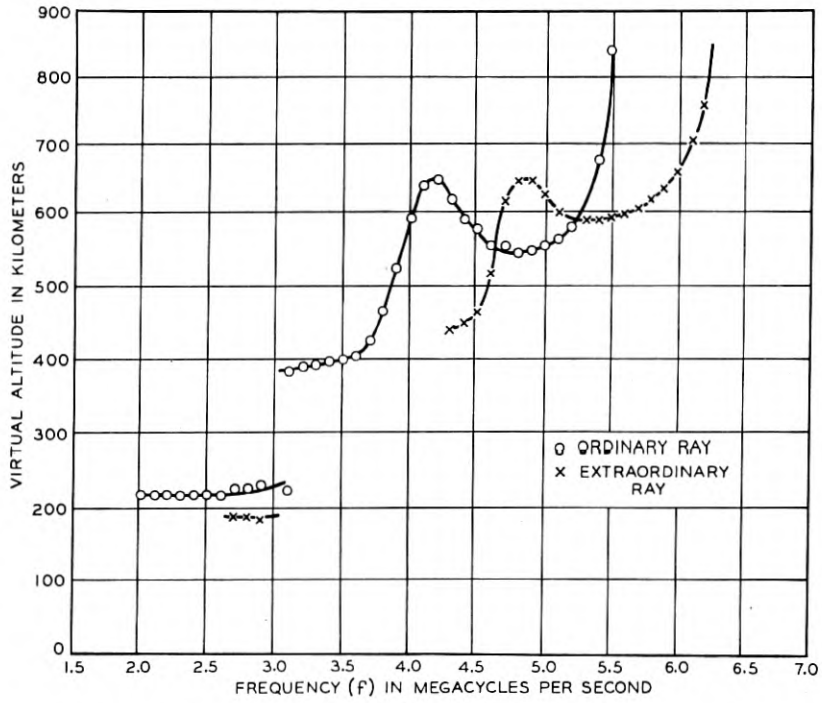

Fig. 4—Example of characteristic or (h', f) curves, showing gap between E-branch and F-branch, and crinkle in F-branch indicating presence of F_1-layer. Duplication of curve due to earth's magnetic field (page 479). (Appleton.)

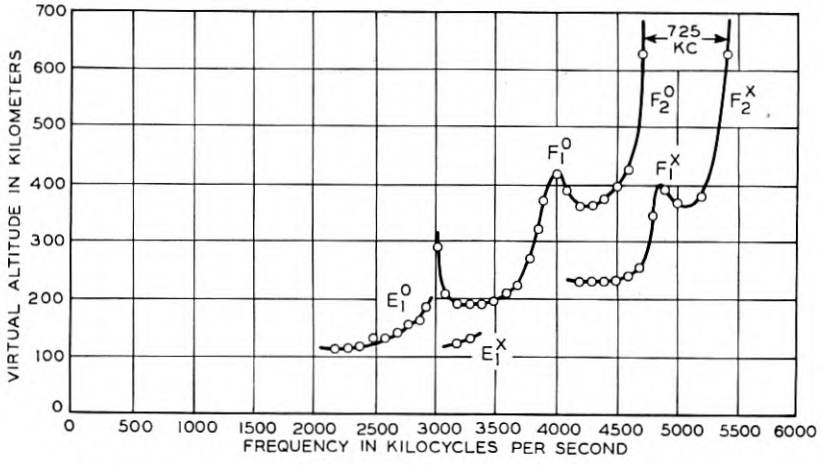

Fig. 5—Another example of (h', f) curve, showing gap and crinkle. (Schafer and Goodall.)

Figures 4 and 5 show two examples of these curves, from data obtained while the sun was high in the sky. Actually there are two curves in each of the figures; the appearance is that of a single curve, repeated with a sidewise shift. I mention that this repetition is due to the earth's magnetic field, but ask the reader to ignore for the present the right-hand curve and fix his attention on the left-hand one. Here he will see the E-branch, the gap, and the F-branch. The

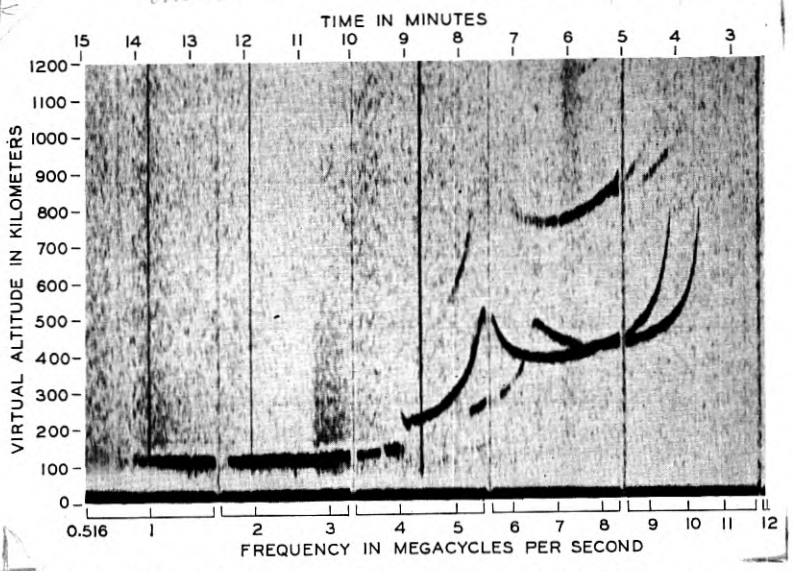

Fig. 6—Characteristic (h', f) curves obtained with the multi-frequency apparatus; sun high in sky, F_1 crinkle apparent. (Carnegie Institution of Washington.)

upturns to right and left of the gap are striking on Fig. 5, insignificant in Fig. 4. The F-branch is deformed by an enormous hump or crinkle. This is supposed to correspond to a second gap, the upturns on right and left being so pronounced as to give a perfect semblance of a peak; indeed one sees in Fig. 5 how readily the gap between E and F might have been drawn as a peak. Curves of this sort are therefore taken as evidence for three layers in the ionosphere, denoted by E and F_1 and F_2. Sometimes there are signs of a fourth, lying between E and F_2, and denoted by M or E_2.

So great is the interest in curves like these, and so much do they vary from time to time and from place to place, that lately there have been more than a score of stations over the world engaged in making them. At some of these the tracing of the curves is speeded up and made incessant by a remarkable machine developed at the

Bureau of Standards and the Carnegie Institution. Automatically sending out the signals ten times in a second, and changing the frequency by (on the average) 1600 cycles between each signal and the next while the photographic film is moved a tiny bit from left to right, this "multi-frequency apparatus" traces the (h', f) curve over the

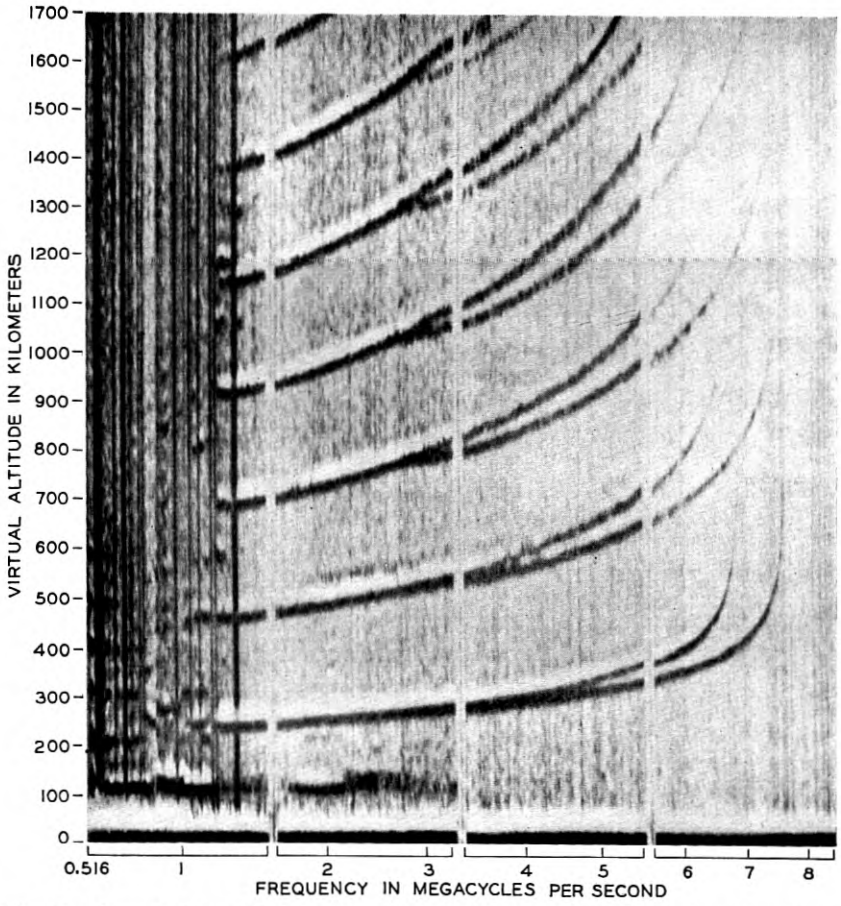

Fig. 7—Characteristic (h', f) curves obtained with the multi-frequency apparatus; sun low in sky, F_1 crinkle missing. (Carnegie Institution of Washington)

frequency-range between 0.516 mc. and 16 mc. in fifteen minutes, and then goes right back and does it over and over again. Figures 6, 7 and 8 show individual curves thus automatically taken, and Fig. 15 a sequence of them spanning several hours of the day.

In Fig. 6 are curves with crinkles in the F-branch, similar to those of Figs. 4 and 5. In Fig. 7, however, the crinkle is missing, and the

F-branch sweeps smoothly and slowly upward from its commencement. (The forking signifies that here are two similar curves lying side by side as in the previous figures, but overlapping so much that over a large part of their course they are not distinct.) Many observations have concurred in showing that the crinkle is present only when the

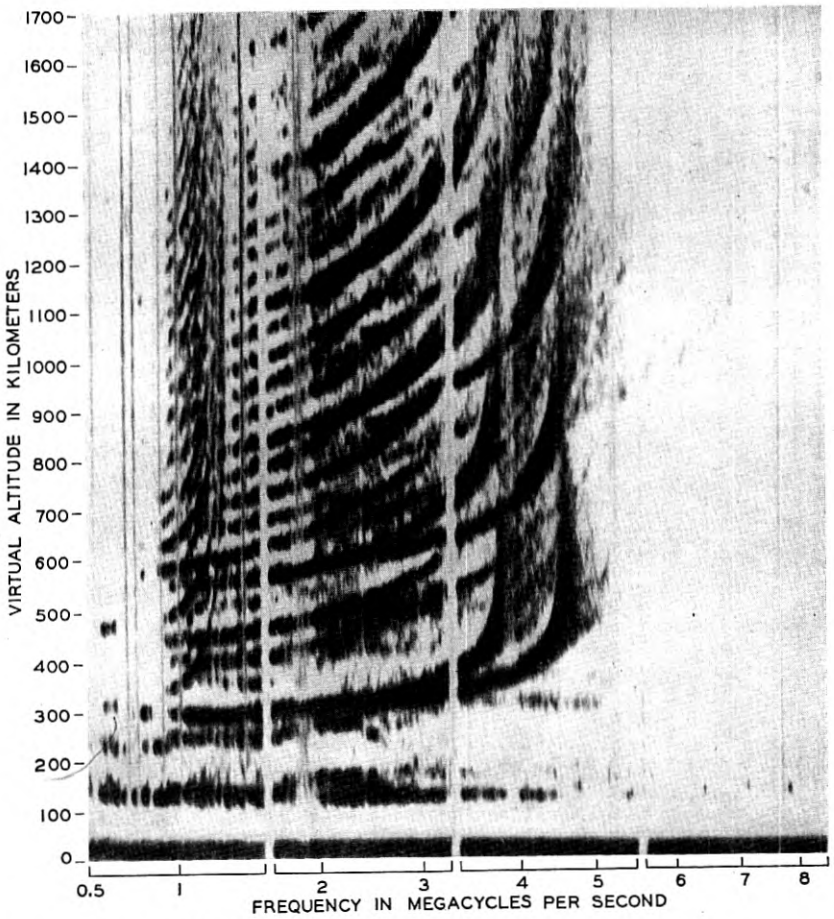

Fig. 8—Multiple echoes. (Carnegie Institution of Washington.)

sun is high in the sky (within some 40° of the zenith)—therefore absent by night and at the beginning and end of day, and indeed absent all day in winter where the latitude is high. This is our first example of the dependence of the ionosphere on sunlight, a very important feature.

In all of these photographs the curves are repeated several times along the vertical direction. This signifies echoes which have traveled four, six, eight or more times between ground and ionosphere, being reflected by both. Figure 8 shows a wonderful multitude of such echoes.

The curves of Fig. 9 are sketches generalized from many data. Contrasting those on the left with those on the right, we see the

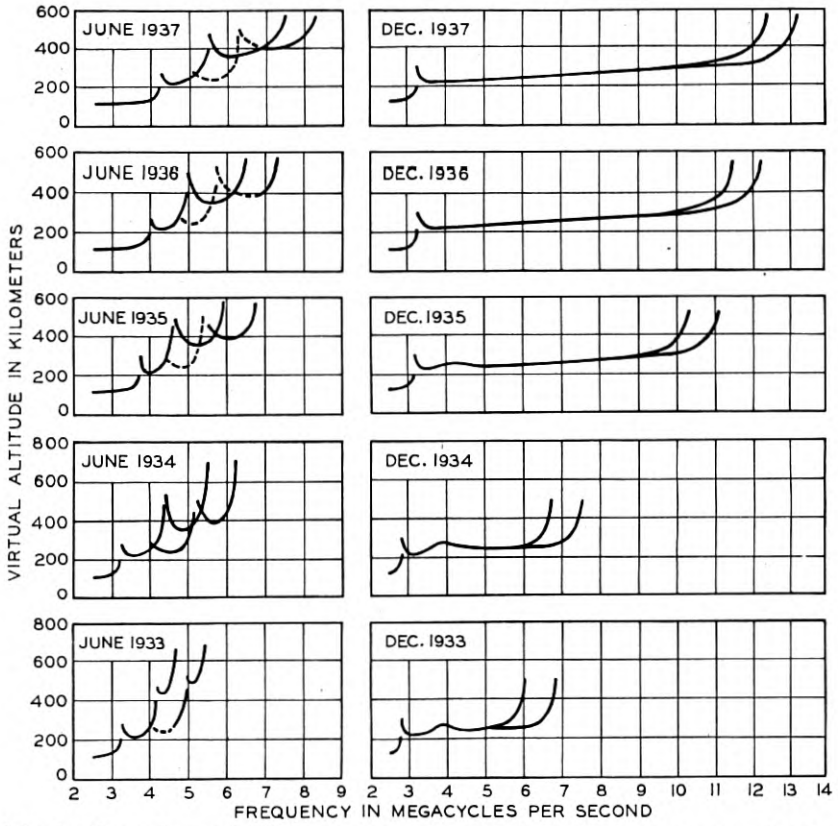

Fig. 9—Dependence of (h', f) curve on season and on the sunspot cycle (Smith, Gilliland and Kirby: National Bureau of Standards).

crinkle prominent in the ones, missing or feeble in the others. This is the difference between summer and winter. All of the data were taken near noon, but though the District of Columbia is not exactly in polar latitudes, the sun in December does not rise far enough in Washington's sky to bring out that feature of the ionosphere of which the crinkle is the sign. Running the eye along either column, one

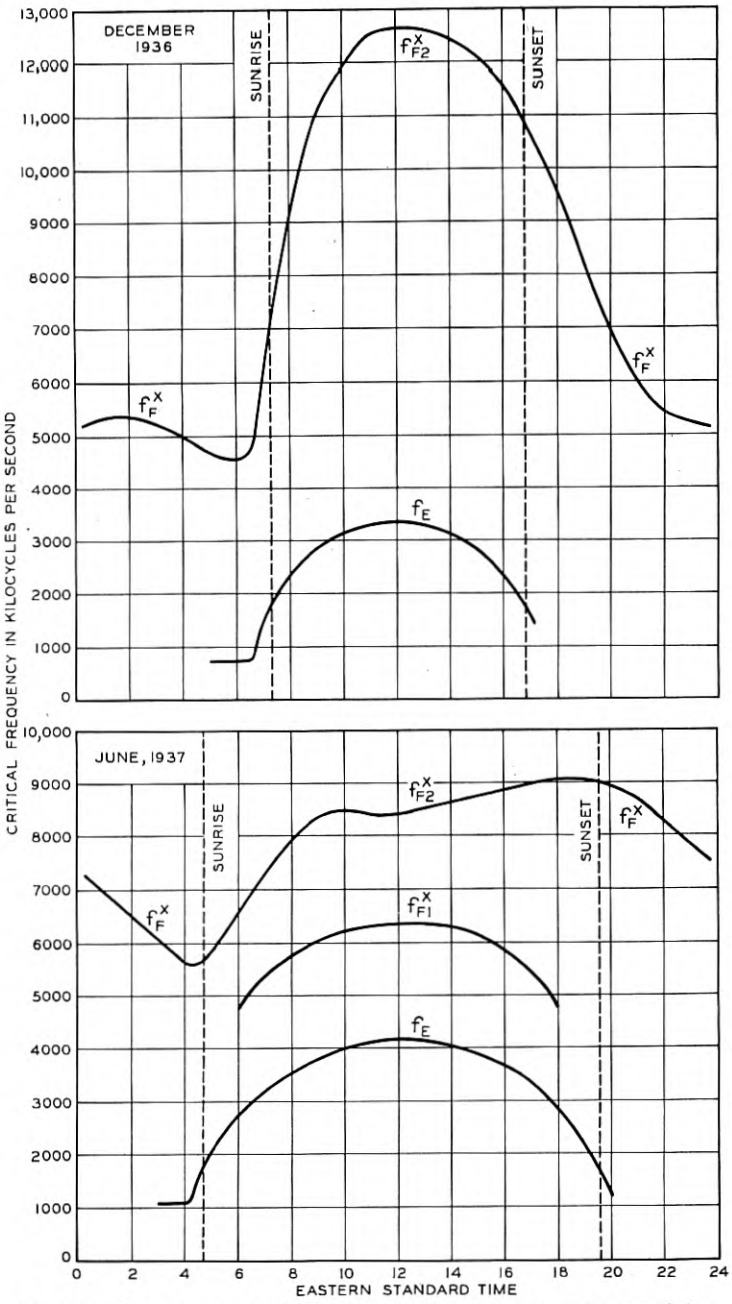

Fig. 10—Dependence of critical frequencies on season and hour of day. (Smith, Gilliland and Kirby.)

sees the curve lengthening out to the right as year follows year. During this series of years the sun spots were growing more frequent: the sun was in the ascending part of that eleven-year cycle of its fever, of which the sun spots are one of the manifestations, while the form of these curves is another. Do not, however, misread this statement as meaning that the curve lengthens out, when and only when there are sun spots on the solar disc! there is no such correlation.

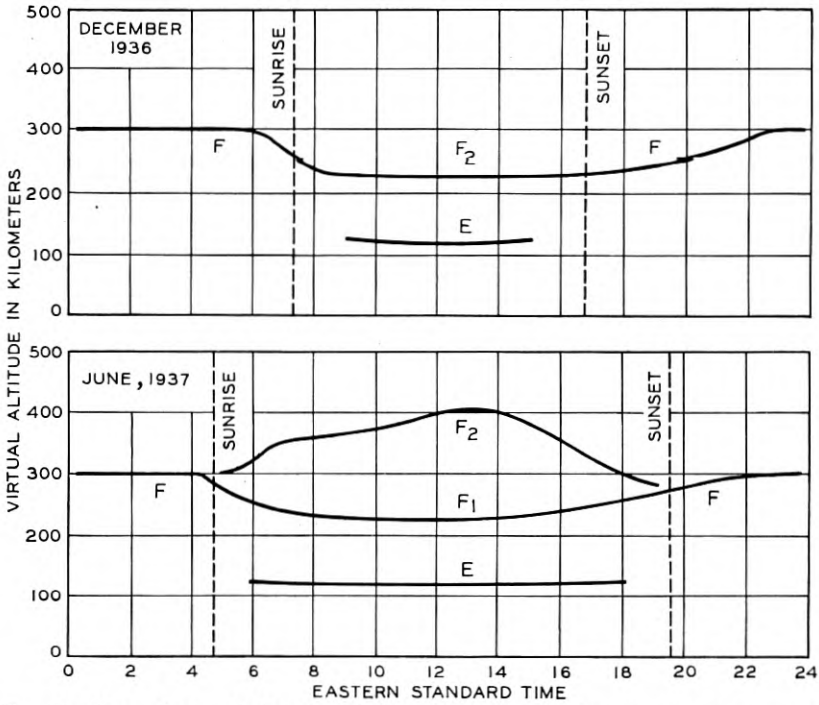

Fig. 11—Dependence of virtual altitude of the crowns of the principal layers on season and hour of day. (Smith, Gilliland and Kirby.)

Another way to study dependence on the sun is to pick out salient features of the (h', f) curves, and see how they vary. Such features are the abscissae and the ordinates of the right-hand ends of the branches and of the top of the F_1 crinkle. The abscissae are the penetration-frequencies for the layers E, F_1 and F_2; the ordinates are the virtual heights (not the true heights) of what I have called the crowns of these layers. All are shown, in their dependence on season and hour of day, in Figs. 10 and 11.

It may have struck the reader that for the last five paragraphs there has been no allusion to the theory, the correlation of the (h', f)

curves with the sun having been presented as if for its own sake. Much work in the field does stop at this point, and is not without value in spite of its stopping there. The theorist indeed may care for an (h', f) curve only as material for deducing the (N, z) curve—the distribution-in-height of the ions—to which he aspires. It is, however, fortunate that this is not the only value of the (h', f) curves, for as we now shall see, the derivation of the (N, z) curves from them is full of difficulties.

Perhaps the greatest of these difficulties springs from the dependence of the signal-speed on N, which is to blame for the difference between virtual height h' and true height h. Even if it is fully justifiable to identify signal-speed with group-speed v, the difficulty is not banished. It resides in the fact that $(h' - h)$ depends not on things already known but on the very thing one is striving to find out, to wit, the distribution of ion-density in the ionosphere. The value of h for any particular h' depends indeed not on the value of N at that height alone, but on the values of N at all inferior heights. The problem is somewhat like having to solve for x an equation in which x appears badly entangled on both sides of the equals sign. The mathematical technique is difficult and approximative.

Another major difficulty resides in the fact that when N varies by an appreciable fraction over a distance equal to a wave-length of the waves, the consequences of the theory become a good deal more complex than those embodied in the simple equations (4) and (5). For instance, partial reflection may occur at a level where N is rising rapidly, though as yet far below the mirror-density N_c. One cannot therefore say that whenever an echo is observed on a frequency f, there *must* somewhere exist an electron-density related to f by (5). The literature is full of allusions to mystifying echoes, some of which are ascribed to partial reflection. Figure 2 shows an E-echo and an F-echo received from the same signal; and many a (h', f) curve shows the E-branch running along for quite a distance underneath the F-branch, instead of stopping at just the abscissa where the F-branch begins.[5] Yet on the other hand, the E-layer may be denser at some places than it is at others of equal altitude, and parts of a signal may be reflected from the places of high density while other parts slip between these and go on to the F-layer. The assumption that N depends on z only, which hitherto has been taken for granted in this paper as it is in most theory, is in fact very assailable; and people are

[5] This is so well-known a phenomenon that lengthy papers have been written about it under the name of "abnormal E-ionization," though it seems too common to deserve the adjective "abnormal."

beginning to study the distribution of N in the horizontal plane, e.g. by using obliquely-sent as well as vertically-emitted signals.

Now we turn briefly to a difficulty affecting not the relation between h' and h, but the relation between N_e and f presented as equation (5). This equation was based on the tacit assumption that the electric force on a single electron is the same as though there were no other electrons at all in the ionosphere. The assumption has been doubted, and quite a polemic has ranged about it. The question is in fact a special case of one of the most pestiferous questions of all mathematical physics, occurring for instance in the theory of magnetized bodies and of bodies polarized electrically: when a great many similar atoms side by side are exposed together to an external field, how is the force suffered by any one of them modified by the presence of its equally-affected neighbors? One strongly-held position is, that there *is* such a modification which manifests itself in a factor 3/2, to be multiplied into the right-hand member of equation (5). A test experiment has been devised, and the early results have favored this theory. The presence or absence of this factor alters in equal proportion all the ordinates, but does not modify in the least the trend of the $N(z)$ curve; but the student specially interested in numerical values of N must discover, from each paper wherein such are given, which formula was used in computing them.

After uttering all these warnings about the theory underlying the (N, z) curves, I will risk a few statements about the curves themselves.

The shape of the (N, z) curve, when the sun is low in the sky and there is no crinkle in the $(h'f)$ curve, is roughly that of Fig. 3*A*. If the sun is within some 40° of the zenith and the crinkle is present in the $(h'f)$ curve, the theory indicates not that the F-peak of Fig. 3*A* has split into two, but rather that a bulge has appeared on the left-hand side of the F-peak. The letters F_1 and F_2 are then applied to the bulge and the peak, respectively. If the shape of the (h', f) curve indicates yet another layer between E and F_1, it appears as a small hump in the valley between the peaks of Fig. 3*A*.

As for the N-values, those of most interest are those corresponding to the crests of the peaks; or to define them better by staying closer to the data, they are the ones corresponding to the points on $(h'f)$ curves which adjoin the gaps or lie at the tops of the crinkles. These may be called the values corresponding to the "crowns" of the several layers.

At Huancayo in the Peruvian Andes, at a typical summer noon, N has the values $1.8 \cdot 10^5 - 3.3 \cdot 10^5 - 1 \cdot 10^6$ at the crowns of the layers E, F_1, F_2: so says Berkner. At Slough near London, at noon

on a certain day of early spring (1933) Appleton found $1.2 \cdot 10^5$ at the crown of E and $3.8 \cdot 10^5$ at the crown of F, the F-layer being at that time and place not differentiated into F_1 and F_2. These figures are not far apart, if E be compared with E and F_1 with F; but with a little search I could have found plenty of values differing much more greatly, as is attested by Figs. 9 and 10. From the former of these we have already deduced that critical frequencies vary as the sunspot cycle proceeds: I now add that from minimum to maximum of the cycle just ending, N at the crown of the E-layer increased by three-fifths while N at the crown of F_2 went up no less than fourfold! From Fig. 10 we infer, by squaring the values of critical frequencies, how great is the change of these N-values with hour of day. Sudden unaccountable changes also occur; one evening over Cambridge (Massachusetts) the N-value for the E-layer was more than tenfold the values given above, being ascertained by Mimno as $2.8 \cdot 10^6$!

I therefore summarize, as precisely as seems justifiable: the N-values at the crowns of the layers vary with hour of day and time of year and year of the sunspot-cycle very markedly, not to speak of sudden unexplained fluctuations; and 10^5 to 10^6 electrons per cc. is a good figure to keep in mind for the order of magnitude thereof.

To terminate this section I show Fig. 12, in which the delay of the echo for a certain frequency (2 mc.) is plotted against time during the hours preceding and following dawn. Interpreting with the aid of Fig. 3*A*: during the night the E-peak was too low to echo back the signals of this frequency, which accordingly climbed farther and found their mirror in F; but at 6:35 A.M. very sharply, the E-peak increased in height to just the extent needed to intercept them. Or since confusion may arise from using the word "height" in two senses, I express what went on in an exacter way: during the night the electron-density at the crown of the E-layer was inferior to the mirror-density for $f = 2$ mc., but with the oncoming of day it rose, and at 6:35 A.M. very sharply it attained and overpassed that mirror-density.

I recall that Fig. 11 exhibits how the virtual altitudes of the layer-crowns vary with hour of day and season of the year in the sky over Washington. It is evident that E is a fixture of the ionosphere with a virtual height surprisingly steady at the close neighborhood of 120 km, while F rises and falls in the course of a winter day, rises and falls and divides and merges again during a day of summer.

Now we must take brief notice of a difficult subject: the forkings and the doublings of the (P', f) curves, and the theory which finds their source in the earth's magnetic field.

A magnetic field, the earth's or any other, should have no effect whatsoever on radio waves so long as these are traveling in air composed entirely of neutral molecules. When, however, the waves are setting electrons into motion, the moving electrons are affected by the field, which has a twisting action on their paths. We have seen already that the moving electrons react, so to speak, upon the waves, raising the wave-speed thereof. By altering the motions of the electrons, the magnetic field will influence at second hand the waves themselves. But will the result be perceptible? In view of the fact

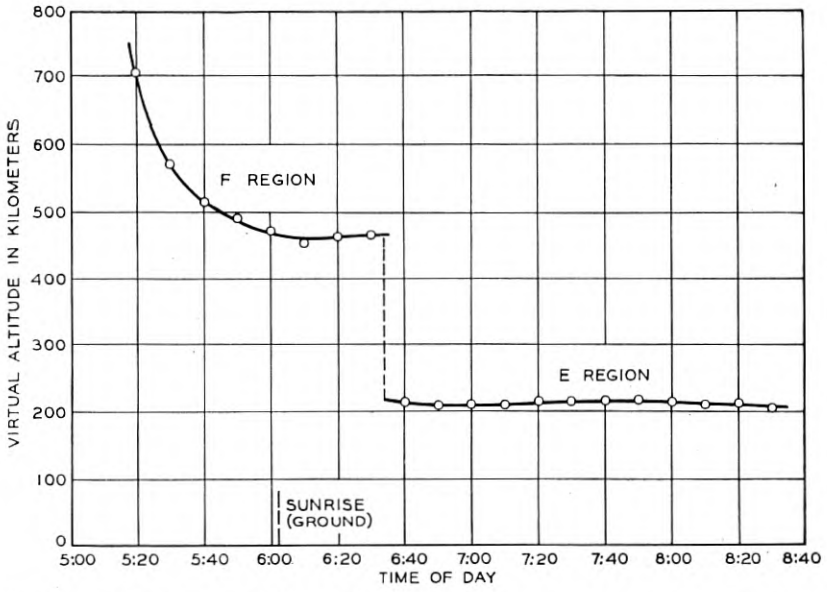

Fig. 12—Ceiling for signal of fixed frequency shifting near sunrise from F to E as the electron-density of the E-layer increases with increase of light. (Appleton.)

that the earth's magnetic field is very feeble by comparison with the fields between the poles of our electromagnets great or small, or even with those around the horseshoe magnets which are playthings, one might well think the influence not worth the trouble of computing. But those who first undertook to compute it—Nichols and Schelleng in America, Appleton independently in England, in the winter of 1924–25—found it a serious influence, and very well worth the trouble.

The problem is one of those which are not very hard to state, but can be very tedious to solve except in special cases which may or may not be of practical importance. For this problem it happens that two of the special cases can be solved with relative ease, and one at

least is realizable in practice. I cannot venture to give the theory of even this one, but at least I will attempt to describe what happens.

The special case occurs when the waves are traveling at right angles to the field. Since they travel vertically,[6] it is necessary to find a place on the earth where the field is horizontal. Such places are found in the equatorial regions only, and these are not precisely crowded with universities or engineering experiment stations. However, the Carnegie Institution of Washington was inspired, several years ago, to set up a station in just such a place: Huancayo, in the Andes of Peru. Here they established long straight horizontal antennae, one running north-and-south, another east-and-west, and yet another northeast-and-southwest. In the waves which mount from these to the ionosphere and then come bouncing back, the electric field $E_0 \sin nt$ —henceforth to be called "the electric vector"—is faithful to the direction of the antenna. They are called "plane-polarized waves."

When the north-south antenna is used, the electrons are impelled to and fro in the north-south direction which is that of the magnetic meridian. Now as is well known, an electron moving parallel to a magnetic field behaves just as it would if there were no such field at all. These waves ought therefore to behave according to the theory which we set up while we were still disregarding the magnetic field. They are the so-called "ordinary waves" or "o-waves."

When the east-west antenna is used, the electrons of the ionosphere are impelled to and fro in the east-west direction, which is transverse to the earth's magnetic field. This is just the condition for the maximum amount of meddling by the field in the motion of the electrons. The meddling consists in bending the electron-paths into curiously twisted arcs. *The action of the magnetic field is tantamount to strengthening the electron-current-density I_e parallel to the electric vector.* It will be recalled (from page 464) that it is I_e which for small N-values cancels a part of the displacement-current and so speeds up the waves, and for a certain critical N-value cancels the whole of the displacement-current and so brings about total reflection. So, for these "extraordinary waves" or "x-waves," a given N-value produces a greater augmentation of the wave-speed, and the critical N-value for total reflection is smaller, than for the ordinary waves. The signal composed of x-waves, mounting into the ionosphere, finds its appropriate mirror at a lesser altitude than does the signal composed of o-waves, and it gets earlier back to earth. It may indeed come

[6] In addition to the other advantages of sending the waves up vertically, there is the feature that the angle between their line of motion and the field is the same when they are going up and when they are returning.

back from E while the o-signal goes on to F or from F_1 while the o-signal goes on to F_2, or from F_2 while the other goes irretrievably forth into space.

When the northwest-southwest antenna is used, the ionosphere takes charge of the signals, and separates them into an o-component and an x-component. Each travels according to its proper law, and the x-component reaches its lower-down mirror earlier and beats the o-component back to earth. Two echoes return instead of one. The earlier is plane-polarized with electric vector east-and-west; the laggard is plane-polarized with electric vector north-and-south. Suppose a long straight horizontal antenna is used to respond to the returning signals. It will respond to both, if pointed north-west-southeast; only to the earlier, if pointed east-and-west; only to the later, if pointed north-and-south.

All the foregoing were statements of theory at first, but thanks to the experiments of Wells and Berkner at Huancayo, they now are statements of data as well.[7] Figure 13 exhibits a small selection from the data.

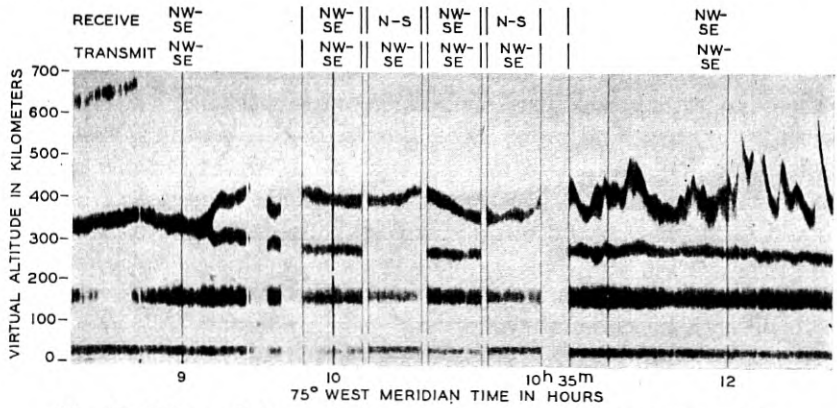

Fig. 13—Echoes of plane-polarized signals near the geomagnetic equator. (Wells and Berkner.)

Now look again at Fig. 4: formerly I asked the reader to ignore one of the curves, but now we will compare the two. The circles and the crosses indicate the o-wave and the x-wave respectively. One is

[7] This is a good place to speak of a question which may already have occurred to many readers, viz. the question why we assume the charged particles in the ionosphere to be free electrons rather than charged atoms or molecules. Were they of atomic or molecular mass, the separation of the o and x echoes would be inappreciable, and the "gyro-frequency" later to be mentioned (page 482) would be quite outside of the radio range. It is not, however, excluded that among the free electrons there may be a great multitude of charged atoms, perhaps even many times more numerous than they, though much less influential.

a fairly close copy of the other, shifted a certain distance along the horizontal axis. From the magnitude of the shift it is possible to compute the strength H of the earth's magnetic field; or let me rather say, it is possible to compute a numerical value which must agree with H, or else the theory will be vitiated. When the computation is made, the value turns out to be just a few per cent less then the field strength at ground-level. The action of the field through the electrons on the waves is exercised only in the ionosphere, which is hundreds of kilometers up in the sky; and it is quite reasonable to believe that these few per cent are actually the falling-off in the field strength from the ground up to that level. Such is the present belief, and many of those who work in terrestrial magnetism are happy over the prospect of measuring thus the field in regions where there seems to be no greater hope of anyone ever actually going, than of going to the moon.

Actually Fig. 4 shows data obtained in England, which is far from the equator; I point this out in order to mention that even when the waves are traveling obliquely to the earth's magnetic field, there is a separation of the signals into pairs of echoes, and these are still amenable to theory. In this general case of oblique transmission, the waves are polarized elliptically—a feature difficult to visualize without a certain amount of specialized knowledge, but lending itself to some very neat and pretty experimental tests.[8] In the special case of transmission parallel to the magnetic field, waves initially plane-polarized should remain of this character but their plane of polarization should rotate as they proceed. There are indeed so many curious and interesting details of the influence of the field through the electrons on the waves, that a writer must be ruthless in ignoring them if he is to observe decent limits of space. I will mention only in closing that the "gyro-frequency" $eH/2\pi mc$—which in our latitudes is around $1.3 \cdot 10^6$ mc.—plays the part of a resonance-frequency. Waves too close to this frequency are liable to great, not to say distressing, anomalies in transmission. Theoretical statements about waves in general are likely to assume two different forms, one appropriate to those of frequency higher and the other to those of frequency lower than the gyro-frequency; it is the former which appears in this paper.

In Fig. 14 there appears something which, if the war had begun before its discovery, would perhaps have been called a "blackout."

[8] It may perhaps be regarded as obvious that the ellipse of polarization should be described in opposite senses in the northern and southern hemispheres, since in one hemisphere the magnetic lines of force are coming out of the ground, in the other they are diving in. This inference was tested by a special experiment in Australia, and the result was taken as establishing the "magneto-ionic theory," as this general theory is often called.

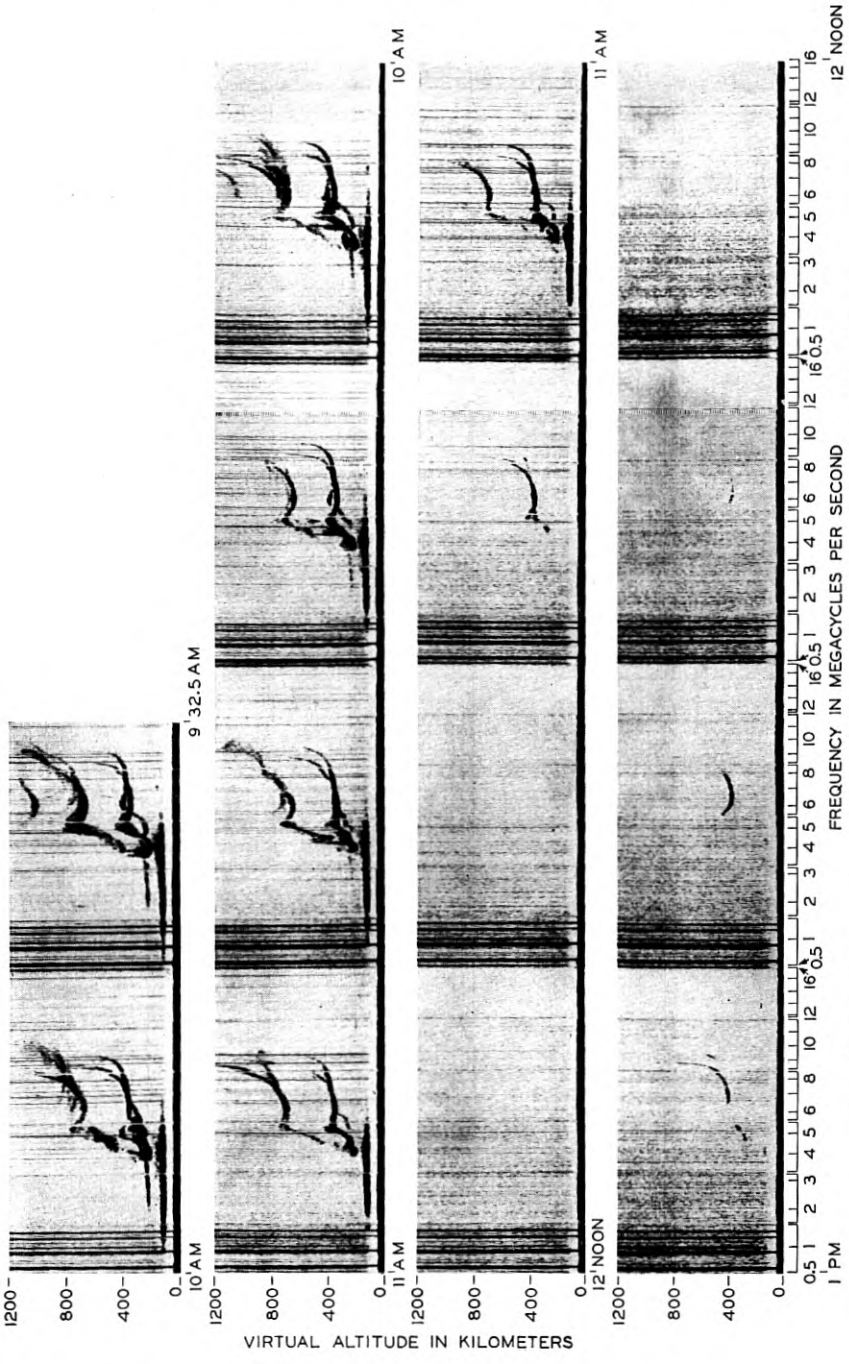

Fig. 14—Progression and recession of a fadeout. (Carnegie Institution of Washington.)

Being discovered however some years before the war, it was and is called a "fadeout." The apparatus was that which I mentioned on page 471; accordingly each of the pictures was traced in fifteen minutes, and as soon as each was finished the next was begun. See how the pattern of (P', f) curves, familiar and sharp in the earlier pictures, dissolves into fragments and then is completely wiped out! Later on it begins to come back piecemeal, and finally is restored as good as ever.

Since attention was focussed on such events in 1935, they have been reported by the scores in every year, varying in duration and in severity. It requires no (h', f) curve to show them, since ordinarily they cut off communication by radio, and with the sharpness of a knife. Many an engineer, to quote from Dellinger, has "dissected his receiving equipment in the vain effort to determine why it suddenly went dead." Over broad areas the extinction is sudden and simultaneous in many fadeouts, more gradual in others; the restoration is as a rule more gradual.

Shall we interpret this strange and striking effect as a sudden vanishing of the ionosphere and all the reflecting layers thereof, or as a swallowing-up of the signals by something which is suddenly created underneath the ionosphere? Against the first suggestion it is to be said, that no one can image anything which might so suddenly frighten all the electrons of the ionosphere back under cover, so to say—drive them into the arms of their parent molecules in a few seconds or minutes—when all day and even at night they manage to hold their freedom. Such a graph as Fig. 15 speaks also against it forcibly. Here in the upper part of the figure we see the critical frequencies [9] of F_2, F_1 and E as located every fifteen minutes on (P', f) curves such as those of Fig. 14. Each flock of data lies along a curve which, intercepted though it is by the fadeout, resumes so nearly at the level where it left off that one can hardly believe that the ionosphere totally vanished in between.[10]

As for the curve marked "f_{MIN}" in Fig. 15, it represents the lowest frequency at which echoes are observed. I have said nothing as yet about there being such a minimum-frequency. How indeed can there be one, and why should signals of any frequency however low fail to be echoed, since the mirror-density for any higher frequency is *a fortiori* more than a mirror-density for any lower?

[9] For the ordinary waves, as indicated by the superscript in symbols such as f^o_E.

[10] In violent magnetic storms the ionosphere is so convulsed that the echoes lose their sharpness entirely, and (h', f) curves like those of Fig. 7 are replaced by broad smudges; or echoes may vanish altogether. These are quite different from fadeouts.

Attempting to supply an answer to this question, I point out that in such part of the theory as I have thus far given, there is nothing corresponding to *absorption.* This is because the electron-current

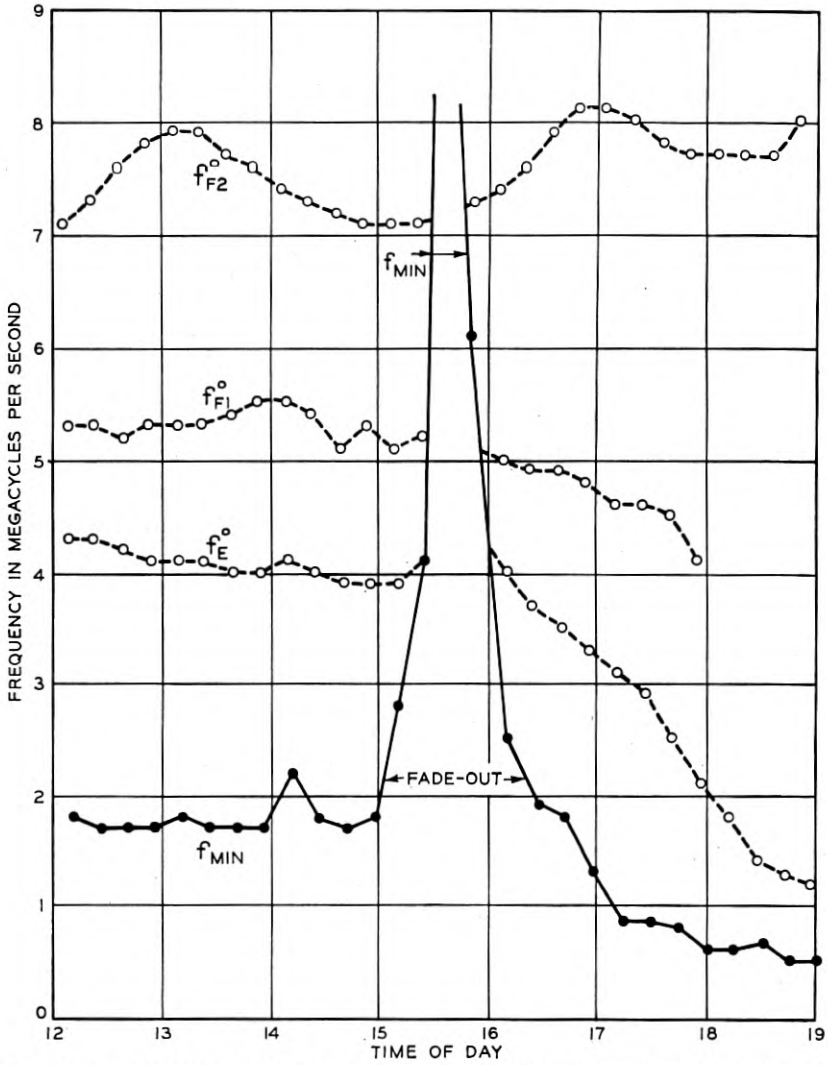

Fig. 15—Trend of the critical frequencies and of the minimum echoed frequency as a fadeout proceeds. (Berkner.)

(I_e in the former notation) is in quadrature with the electric vector in the waves, being thus a "wattless current." This in turn is because we have assumed each electron to be entirely free, oscillating in the

wave-swept aether as though there were nothing else in the world but itself and the waves. If, however, the electron were occasionally to strike and bounce off a molecule it would leave some (even though but a small part) of its kinetic energy behind, and this would have to be replenished by the waves. Indeed as soon as collisions are taken into account, the algebra tells us that the electron-current is no longer in perfect quadrature with the electric vector. Dissipation of energy occurs together with reflection, and if it is sufficiently great —if, that is to say, the electrons collide often enough with molecules— it takes the place of reflection. In the terms of my acoustic simile on an earlier page, the signals are no longer echoed back from the hard slopes of the mountain-range of Fig. 3*A*, but are swallowed up and lost as if in something soft and woolly.

One would expect absorption to occur in the lower reaches of the ionosphere rather than in the upper, since the air is denser there and the electrons suffer many more collisions. The absorbing layer, that is to say, must be situated just where it is able to cut us off from the reflecting layers by dissipating the signals which we send. Why should it do so occasionally with such completeness, and most of the time not do so at all, for any except the lowest frequencies?

For an answer to this query, one looks again to the sun. Ordinarily, we will suppose, the ionizing agent coming from the sun penetrates deep enough into the air to form the reflecting layers high overhead, but is nearly consumed in so doing. Occasionally, though, the sun sends forth a quite abnormal transitory burst of radiation, so strangely constituted that it passes the reflecting layers without contributing to them or weakening itself, and continues so far down that at the level where it at last engenders free electrons they constitute a layer absorptive and not reflective. This would be no more than an *ad hoc* assumption, were it not that brilliant eruptions are frequently seen on the face of the sun at the moments when fadeouts are commencing. To some extent it is still an *ad hoc* assumption, for the light whereby the eruption is seen is certainly not ionizing light, and we must assume that the visible light is attended by rays of other wave-lengths having just the properties desired. Coincidence of fadeout and eruption is, however, so frequent, that it would now take a very sceptical mind to reject the assumption. The trend of the curve "f_{MIN}" in Fig. 15 sustains it.

And so I have now come back to the theory that it is radiation from the sun which makes the upper atmosphere into an ionosphere, by detaching electrons incessantly from the molecules thereof. (The detaching must be incessant, for the electrons are always liable to

recapture by the molecules.) Of this theory it may be said that the major facts confirm it, though at night the *E*-ionization persists so tenaciously that we are obliged to seek for a separate agent (meteors, perhaps?); while numerous minor discrepancies can be explained away by making special assumptions which can neither be confirmed nor refuted because there is so little independent knowledge of the upper atmosphere. Not a very satisfactory situation for the present, but at any rate one which offers endless promise!

Thus it can readily be seen that as the ionizing light descends from heaven through the upper air, the ionization per unit volume should at first increase (because the air is getting denser) and then decrease (because the light is getting to be used up). This offers an explanation for a layer; and the mathematical working-out of the idea—due in the main to Chapman—shows that not only the existence but the shape of either peak in the curve of Fig. 3A is compatible with the theory. But there are several layers and peaks, not just one; how does this come about? Well, the atmosphere is a mixture of several gases, differently susceptible to the ionizing light; one can attribute a peak to each gas (indeed more than one to a gas, by invoking different states of the molecules). The height of a peak, the *N*-value at the crown of a layer, should rise and fall as the sun rises and sinks in the sky. This is true of the layers *E* and F_1, as we saw from Fig. 11, and again there is a quantitative theory by Chapman, which is borne out in some though not in full detail. Of F_2 it is not always true, as Fig. 11 proclaims; there is a minimum at noon in summer, and the highest *N*-values of all are attained in winter! One tries to cope with the discrepancy by assuming that as the sun climbs higher in the sky, the F_2 region expands so much in the heat that although the total number in the region is properly increasing, the number in unit volume suffers a decline. The layers do not disappear at night, though the *N*-values shrink. There seems to be plenty of time for recapture of all the electrons between sunset and sunrise, and one is driven to hunt for other causes of ionization which emerge when the sunlight is gone. These remain mysterious. Inrush of meteors into the high atmosphere has been suggested as one of the causes, and also incessant streams of charged particles similar to those which become intense during magnetic storms.

Sunlight is therefore not the only, yet apparently the major factor in maintaining the ionosphere. Not, however, any sunlight that we ever feel! This portion of the sun's outpourings is so thoroughly consumed above that it never reaches down to the levels where we live. Were it not so consumed, we should not be able to communicate by

radio very far over the earth. The reader may think that this is not very important: our ancestors lived without radio, why should we worry about lacking it? Well, it is probably quite true that if the ionosphere were not overhead, we should not be worrying about the lack of radio. We should in fact probably not be worrying about anything at all, for we should not be here to worry. The ultra-violet light of the sun, pouring down upon the surface of the globe unhindered, would work changes so severe on organisms as we know them that life would have to be very different, and perhaps impossible. This lethal light is like an enemy, which in attacking a city spends itself in throwing up a barrier against itself; and the barrier not only keeps the enemy out, but is serviceable otherwise to the dwellers in the city.

Abstracts of Technical Articles by Bell System Authors

Stereophonic Reproduction from Film.[1] HARVEY FLETCHER. On April 9 and 10, 1940, demonstrations of the stereophonic reproduction of music and speech, described in this article, were given at Carnegie Hall, New York, N. Y. These demonstrations represented the latest development in a series of researches by Bell Telephone Laboratories, the first step of which was demonstrated in 1933 when a symphony concert, produced in Philadelphia, was transmitted over telephone wires to Washington, and there reproduced stereophonically and with enhancement before the National Academy of Sciences.

For the present demonstrations, original recordings of orchestra, choir, and drama were made at Philadelphia and Salt Lake City; and at a later audition the artist or director was able to vary the recorded volume and to change the tonal color of the music to suit his taste. At will, he could soften it to the faintest pianissimo or amplify it to a volume ten times that of any orchestra without altering its tone quality, or he might augment or reduce the high or low pitches independently. The music or drama so enhanced is then re-recorded on film, with the result that upon reproduction, a musical interpretation is possible that would be beyond the power of an original orchestra, speaker, or singer to produce.

Wave Shape of 30- and 60-Phase Rectifier Groups.[2] O. K. MARTI and T. A. TAYLOR. The installation of mercury arc rectifiers with a total capacity of 82,500 kilowatts by the Aluminum Company of America at Alcoa, Tennessee and Massena, New York, was accompanied by widespread increases in the inductive influence of the interconnected power supply networks with resultant increases in the noise on exposed telephone circuits. Because of the size of these installations, and the complexity of the supply systems, it appeared impracticable to limit the rectifier harmonics by the use of frequency-selective devices, which have been successfully applied to certain smaller installations. However, the results of a cooperative study indicated that by means of a relatively simple arrangement of phase shifting transformers, the equivalent of 30- or 60-phase operation of the rectifier stations could be secured. In this way, the important harmonic components on the power systems were reduced to relatively small values, and wave shape and noise conditions were restored practically to normal.

[1] *Jour. S. M. P. E.*, June 1940.
[2] *Elec. Engg.*, April 1940.

This paper describes briefly the voltage and current relations, preliminary tests on a small-scale rectifier, the phase shifting transformers and their application to this particular situation, and presents data to indicate their effectiveness.

A New Quartz-Crystal Plate, Designated the GT, which Produces a Very Constant Frequency over a Wide Temperature Range.[3] W. P. MASON. In this paper, a new quartz-crystal plate, designated the GT, is described which produces a very constant frequency over a wide temperature range. This crystal does not change by more than one part in a million over a 100-degree centigrade range of temperature. This crystal obtains its great temperature stability from the fact that both the first and second derivatives of the frequency by the temperature are zero. Both the frequency and the temperature coefficient can be independently adjusted.

This crystal has been applied in frequency standards, in very precise oscillators, and in filters subject to large temperature variations. It has given a constancy of frequency considerably in excess of that obtained by any other crystal. A crystal chronometer, using this type of crystal, was recently lent to the Geophysical Union for measurements on the variation of gravity and the chronometer is reported to have kept time within several parts in 10 million, although no temperature control was used.

Room Noise at Telephone Locations—II.[4] D. F. SEACORD. Room-noise data, based primarily on measurements made at about 900 locations in and around Philadelphia and Chicago under winter conditions, were reported informally to the conference on sound at the 1939 A. I. E. E. winter convention. The present article supplements the earlier material and includes a summary of room-noise conditions expressed in terms of annual averages based on both the winter survey data previously discussed and the summer survey data that had been obtained at about 1,300 locations but had not been completely analyzed at the time of the earlier report. The summer survey included 500 measurements at locations previously measured during the winter in and around Philadelphia and Chicago, and 800 measurements at locations in and around Cleveland, New York City, northern New Jersey, and Philadelphia. Annual average as used in this article is the mean of winter and summer measurements. In addition, the present article includes a brief discussion of outdoor noise and the relative frequency of occurrence of several predominant sources of room noise.

[3] *Proc. I. R. E.*, May 1940.
[4] *Elec. Engg.*, June 1940.

The noise measurements were made with equipment conforming to the specifications described in the A. S. A. "Tentative Standards for Sound Level Meters" (Z24.3–1936), using the 40-decibel loudness-weighting network. The measurements are expressed in terms of sound level in decibels above reference sound level, that is, 10^{-16} watt per square centimeter at 1,000 cycles in a free progressive wave, each measurement being based on the average of 50 individual readings.

Electrical Conductance Measurements of Water Extracts of Textiles.[5] A. C. WALKER. It has been shown that the electrical properties of textiles depend upon chemical composition, water-soluble electrolytic impurities, moisture content and manner of drying the material from the wet state. Selection of a textile for electrical puposes should include consideration of the influence of chemical composition upon the properties of the material, absence of significant amounts of electrolytes, and the method of drying the material from the wet state.

This paper discusses the water extract conductivity method, its correlation with insulation resistance data, and describes a simple, durable electrolytic cell which is convenient to use for the conductivity measurements.

[5] *A. S. T. M. Proc.*, Vol. 39, 1939.

Contributors to this Issue

H. W. BODE, A.B., Ohio State University, 1924; M.A., 1926; Ph.D., Columbia University, 1935. Bell Telephone Laboratories, 1926–. As Consultant in Network Theory, Dr. Bode is engaged in research studies of electrical networks and their applications to various transmission problems.

R. P. BOOTH, S.B. in Electrical Engineering, Massachusetts Institute of Technology, 1925. American Telephone and Telegraph Company, Department of Development and Research, 1925–34; Bell Telephone Laboratories, 1934–. Mr. Booth has been active in the development of line design methods suitable from the interference standpoint for carrier and broad-band transmission.

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–. As Research Physicist, Dr. Darrow has been engaged largely in writing on various fields of physics and the allied sciences.

FREDERICK J. GIVEN, S.B., Harvard and Massachusetts Institute of Technology, 1919. Western Electric Company, Engineering Department, 1919–1925; Bell Telephone Laboratories, 1925–. Mr. Given has been engaged in design and development of transmission apparatus, including retardation coils, condensers, and transformers as well as loading coils and cases.

K. E. GOULD, B.S. in Electrical Engineering, Oklahoma A. and M. College, 1924; M.S. 1925, Sc.D. 1927, Massachusetts Institute of Technology. American Telephone and Telegraph Company, 1927–34; Bell Telephone Laboratories, 1934–. Dr. Gould, formerly engaged in inductive coordination studies, is concerned with transmission measurements at high frequencies.

VICTOR E. LEGG, B.A., 1920, M.S. 1922, University of Michigan. Research Department, Detroit Edison Company, 1920–21; Bell Telephone Laboratories, 1922–. Mr. Legg has been engaged in the development of magnetic materials and in their applications, particularly for the continuous loading of cables, and for compressed dust cores.

TODOS M. ODARENKO, University of Technique in Prague, E.E., 1928. New York Telephone Company, 1928–30; Bell Telephone Laboratories, 1930–. Mr. Odarenko has been engaged in the measurement and study of transmission characteristics of existing and newly developed types of transmission lines.

T. SLONCZEWSKI, B.S. in Electrical Engineering, Cooper Union Institute of Technology, 1926. Bell Telephone Laboratories, 1926–. Mr. Slonczewski has been engaged in the development of electrical measuring apparatus.