599820

Kansas City
Public Library

THE SCHOOL DISTRICT OF KANSAS CITY, MISSOURI

This Volume is for

REFERENCE USE ONLY

7-27—6m—P

PUBLIC LIBRARY
KANSAS CITY
MO

SEP 6 47

From the collection of the



San Francisco, California
2008

PUBLIC LIBRARY
KANSAS CITY
MO

THE BELL SYSTEM TECHNICAL JOURNAL

A JOURNAL DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

EDITORIAL BOARD

J. J. Carty — Bancroft Gherardi — F. B. Jewett
E. B. Craft — L. F. Morehouse — O. B. Blackwell
H. P. Charlesworth — E. H. Colpitts — H. D. Arnold
R. W. King—*Editor* — J. O. Perrine—*Asst. Editor*

VOLUME VII
1928

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

PUBLIC LIBRARY
KANSAS CITY
MO

Bound
Periodical
599820

NOV 21 28

The Bell System Technical Journal

January, 1928

The Measurement of Acoustic Impedance and the Absorption Coefficient of Porous Materials

By E. C. WENTE and E. H. BEDELL

SYNOPSIS: Various ways of determining the acoustic impedance and the absorption coefficient of porous materials from measurements on the standing waves in tubes are discussed. In all cases the material under investigation is placed at one end of the tube and the sound is introduced at the other end. Values of the coefficient of absorption of a number of commonly used damping materials as obtained by one of the methods are given. Several types of built-up structures are shown to have a greater absorption coefficient for low frequency sound waves than is conveniently obtainable by a single layer of material.

THE most commonly used method of determining the sound absorption coefficient of a material is that devised by the late Professor W. C. Sabine. In this method the reverberation time of a room is measured before and after the introduction of a definite amount of the material. This method has the great merit that the values so determined usually apply to the materials precisely as they are ordinarily used in rooms for damping purposes. However, it is tedious and requires a very quiet room and large samples of the materials. A simpler scheme has been devised by H. O. Taylor,[1] in which the absorbing material is placed at one end of a tube. The coefficient of absorption is determined from a measurement of the ratio of maximum to minimum pressures of the standing waves within the tube when sound is introduced at the open end. Thus only a small sample of the material is required and with suitable apparatus the measurements can be made with great facility. In this paper several modifications of Taylor's tube method are discussed; in addition, it is shown that by a similar method it is possible to determine not only the absorption coefficient but also the acoustic impedance, a quantity which is playing an important part in present day applied acoustics.

GENERAL THEORY

Consider a tube of length l, which is filled with a medium having a propagation constant $P = \alpha + i\beta$ and a characteristic acoustic im-

[1] *Phys. Rev.*, II, 1913, p. 270.

pedance [2] equal to Z_0 per unit area. At one end, O, let the velocity be uniform over the whole cross-section and equal to $\dot{\xi}_1 e^{i\omega t}$. At a distance l from O let the tube be terminated by the material which is to be investigated, and the acoustic impedance of which may be rep-

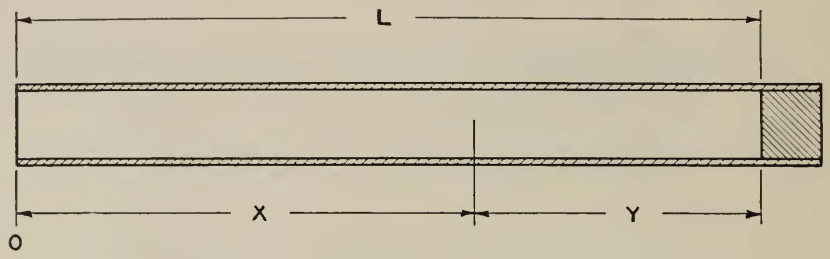


Fig. 1

resented by $Z_2 = R_2 + iX_2$ per unit area. Under these conditions, the pressure, p, at any point in the tube at a distance x from O, is by analogy with the electrical transmission line

$$p = \dot{\xi}_1 e^{i\omega t} \left[\frac{Z_2 \cosh Pl + Z_0 \sinh Pl}{Z_0 \cosh Pl + Z_2 \sinh Pl} \cosh Px - \sinh Px \right] Z_0.^{3} \quad (1)$$

If there is no attenuation along the tube, we get, on dropping the time factor,

$$p = R\dot{\xi}_1 \left[\frac{Z_2 \cos \beta l + iR \sin \beta l}{R \cos \beta l + iZ_2 \sin \beta l} \cos \beta x - i \sin \beta x \right], \quad (2)$$

where $R = c\rho$, the product of the velocity of propagation along the tube and the density of the medium, and

$$\beta = \frac{2\pi f}{c}.$$

Equation (2) indicates numerous possible ways of determining Z_2, e.g., from the values of $\dot{\xi}_1$ and of p at any point in the tube; from the pressures for two values of either x or l, if $\dot{\xi}_1$ is constant; from the pressures at any point in the tube for the unknown and for a known value of Z_2; from the magnitude of p as a function of either x or l. However, we shall confine our discussion to three methods, which appear to be most practicable.

[2] The term acoustic impedance as here used may be defined as the ratio of pressure to volume velocity; the characteristic impedance is this impedance if the tube were of infinite length.

[3] J. A. Fleming, "Propagation of Electric Currents in Telephone and Telegraph Conductors," page 98; 3d Ed.

(a) *Pressure Measured at Two Points in the Tube*

It has already been pointed out that the impedance Z_2 can be determined if the relative phase and magnitude of the pressures at any two points in the tube are known. However, from the standpoint of convenience and precision it appears best to measure the pressures at the reflecting surface and at a point a quarter of a wave-length away. We then have at the reflecting surface $x = l$ and

$$p_2 = R\xi_1 \left[\frac{R_2 + iX_2}{R \cos \beta l + iZ_2 \sin \beta l} \right],$$

and for the point $x = l - \frac{\lambda}{2} = l - \frac{\pi}{2\beta}$,

$$p_1 = R\xi_1 \left[\frac{iR}{R \cos \beta l + iZ_2 \sin \beta l} \right],$$

so that

$$\frac{p_2}{p_1} = \frac{X_2 - iR_2}{R} \equiv Ae^{i\varphi}.$$

Hence

$$\left.\begin{aligned} R_2 &= -AR \sin \varphi, \\ X_2 &= AR \cos \varphi, \\ |Z_2| &= AR. \end{aligned}\right\} \tag{3}$$

If the coefficient of reflection is expressed as [4]

$$Ce^{i\Psi} = \frac{Z_2 - R}{Z_2 + R}, \tag{4}$$

we get

$$C = \left[\frac{1 + 2A \sin \varphi + A^2}{1 - 2A \sin \varphi + A^2} \right]^{1/2},$$

where

$$\varphi = \tan^{-1} \frac{2A \cos \varphi}{A^2 + 1}. \tag{5}$$

The absorption coefficient, which is generally defined as the ratio of absorbed to incident power, is equal to $1 - |C|^2$.

(b) *Tube of Constant Length; the Absolute Value of the Pressure Measured at Points along the Tube*

The method discussed under this section is that adopted by H. O. Taylor for measuring the absorption coefficient of porous materials.

[4] I. B. Crandall, "Theory of Vibrating Systems and Sound," page 168.

For the absolute value of the pressure at any point in the tube we get from equation (2)

$$|p| = \left[\frac{R_2^2 + X_2^2 + R^2 + (R_2^2 + X_2^2 - R^2)\cos 2\beta y + 2X_2 R \sin 2\beta y}{R_2^2 + X_2^2 + R^2 - (R_2^2 + X_2^2 - R^2)\cos 2\beta l - 2X_2 R \sin 2\beta l} \right]^{1/2} R\xi_1, \quad (6)$$

where $y = l - x$.

$|p|$ has maximum or minimum values when

$$\tan 2\beta y = \frac{2X_2 R}{X_2^2 + R_2^2 - R^2}; \quad (7)$$

for the maximum value $2\beta y$ lies in the first and for the minimum, in the third quadrant. We therefore get

$$\frac{|p|_{\max}}{|p|_{\min}} = \left[\frac{X_2^2+R_2^2+R^2+\sqrt{(X_2^2+R_2^2-R^2)^2+4X_2^2R^2}}{X_2^2+R_2^2+R^2-\sqrt{(X_2^2+R_2^2-R^2)^2+4X_2^2R^2}} \right]^{1/2} \equiv A. \quad (8)$$

Let y_1 be the value of y for which the pressure is a maximum; we then have from (7) and (8) and (4)

$$R_2 = \frac{2AR}{(A^2 + 1) - (A^2 - 1)\cos 2\beta y_1}, \quad (9)$$

$$X_2 = \frac{R(A^2 - 1)\sin 2\beta y_1}{(A^2 + 1) - (A^2 - 1)\cos 2\beta y}, \quad (10)$$

$$C_1 = \frac{A - 1}{A + 1}, \quad (11)$$

$$\Psi = 2\beta y_1.$$

The relation (11) can be derived more simply on the classical theory, as it was done by H. O. Taylor. A derivation of (11) is given by Eckhardt and Chrisler,[5] which differs from that of H. O. Taylor. From their derivation it would appear that for (11) to be valid the length of the tube should be adjusted for resonance and that the change in phase at the reflecting surface should be small. The derivation here given shows that (11) is general; it implies only that the waves be plane and that there be no dissipation of power along the tube.

[5] Scientific Paper of the Bureau of Standards, No. 526, page 56.

(*c*) *Tube of Variable Length. Pressure Measured at the Source*

The absolute value of the pressure at the driving end of the tube according to (2) is

$$|p_1| = \left[\frac{R_2^2 + X_2^2 + R^2 + (R_2^2 + X_2^2 - R^2)\cos 2\beta l + 2X_2R \sin 2\beta l}{R_2^2 + X_2^2 + R^2 + (R_2^2 + X_2^2 - R^2)\cos 2\beta l - 2X_2R \sin 2\beta l}\right]^{1/2} R\dot{\xi}_1$$

and $|p_1|$ is a maximum or a minimum when

$$\tan 2\beta l = \frac{2X_2R}{R_2^2 + X_2^2 - R^2}.$$

For the maximum value $2\beta l$ lies in the first and for the minimum, in the third quadrant. We therefore have

$$\frac{|p_1|_{\max}}{|p_1|_{\min}} = \frac{X_2^2 + R_2^2 + R^2 + \sqrt{(X_2^2 + R_2^2 - R^2)^2 + 4X_2^2R^2}}{X_2^2 + R_2^2 + R^2 - \sqrt{(X_2^2 + R_2^2 - R^2)^2 + 4X_2^2R^2}} \equiv A.$$

By analogy from the equations derived in section (*b*) above, we see that

$$R_2 = \frac{2\sqrt{A}R}{(A+1) - (A-1)\cos 2\beta l_1},$$

$$X_2 = \frac{R(A-1)\sin 2\beta l_1}{(A+1) - (A-1)\cos 2\beta l_1},$$

$$C = \frac{\sqrt{A} - 1}{\sqrt{A} + 1},$$

$$\Psi = 2\beta l_1,$$

where l_1 is the length of the tube when p_1 has a maximum value.

Discussion of the Precision of the Methods

Of the three methods of measuring impedance discussed above, the first is undoubtedly the simplest and most convenient, if an a.c. potentiometer is available. Theoretically, in this case the impedance may be determined with a high degree of precision. However, the method presupposes that the points where the pressures are measured are exactly a quarter of a wave-length apart; a more detailed analysis shows that, if A is small, variations in this distance will have a large effect on both the ratio of the pressures and their phase difference. It therefore is necessary to keep the temperature of the tube accurately constant or else to determine the distance corresponding to a quarter

of a wave-length before each measurement. A precise determination of the point a quarter of a wave-length from the reflecting surface may be made by placing a smooth metal block at the reflecting end and finding then the position in the tube at which the pressure is a minimum.

In the other two methods it is relatively less important that the temperature be maintained constant, for the ratio of pressures is affected very little by any temperature variations. In the third method, where the length of the tube is varied, the expressions for R_2 and X_2 are the same as in (*b*), except that in place of the ratio of pressures they involve the square root of this ratio. For small values of pressure ratios the precision is therefore somewhat greater. However, for high values of reflection the ratio becomes very large and great care is required in the experimental set up to prevent errors creeping into the measurements through extraneous vibrations and stray electromotive forces in the measuring circuit. The main advantage of the method in which the pressure at the source only is measured is that a short length of exploring tube is required. If measurements down to a frequency of 60 cycles are made, the tube length must be at least 8 feet. An exploring tube reaching the whole length would ordinarily introduce too much attenuation if it were of sufficiently small bore to prevent resonance effects at the lower frequencies.

Experimental Procedure

In the case of the experimental results here reported the measurements were all made by the method outlined in section (*c*), i.e., the pressures were measured at the source while the length of the tube was varied. The experimental set up is shown in Fig. 2. A piece of Shelby

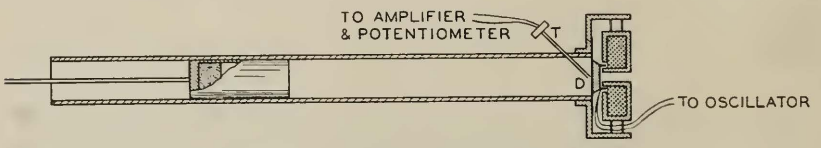


Fig. 2—Diagram of apparatus

steel tubing, 9 feet long, of 3″ internal diameter, and with 1/4″ wall, was fitted with a piston carrying the absorbing material. This piston was made up of a brass tube one foot long with a wall 1/64″ thick, the far end of which was closed with a one-inch brass block. To insure the propagation of plane waves and a constant velocity at the source, the diaphragm at *D* had a diameter of 2⅞″, and a mass of about 100 grams. This was driven with a coil 2″ in diameter situated in a radial magnetic field. The annular gap between the edge of the diaphragm

and the interior of the tube was closed by a flexible piece of leather. To prevent vibrations of the magnet from getting to the tube, the magnet was held in position by flexible supports. The exploring tube t was about 5″ long with a 1/16″ bore which led to the transmitter, T. The voltages generated by the transmitter were measured with an amplifier and an a.c. potentiometer. The potentiometer was used because with it small voltages can be measured and errors due to harmonics are avoided. The proper functioning of the apparatus was determined by measuring the coefficient of reflection with no absorbing material in the piston. Theoretically the reflection should then be practically 100 per cent. The pressure ratios that were actually observed were of the order of 12,000 which corresponds to a reflection coefficient of 98 per cent. Evidently some extraneous pressures or voltages were still present. However, no attempt was made to reduce these further as the materials tested had a reflection coefficient considerably less than this value.

Experimental Results

A brief study was made of the absorption of hair felt, as there is an appreciable variation in the data given by various investigators on the absorption frequency characteristic of felts of presumably the same

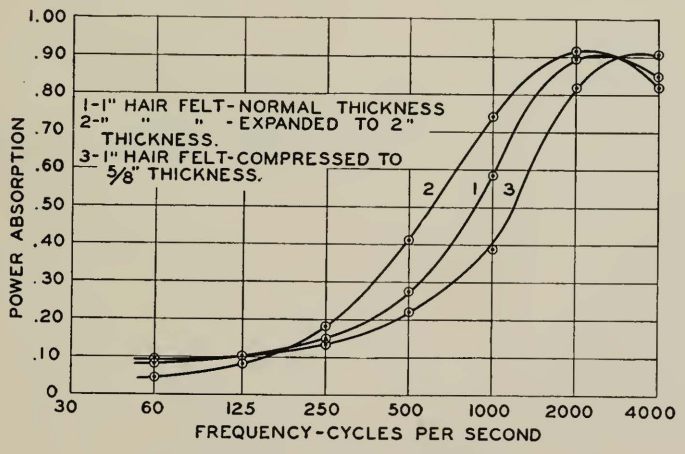


Fig. 3—Power absorbed by hair felt

type. After measurements on several samples it was evident that concordant results could not be expected as the absorption varied considerably with the packing of the felt. This point is illustrated by the curves shown in Fig. 3. These curves were all obtained on the same

piece of hair felt but with different degrees of packing. It is thus evident that a felt which has become loosened by handling may have an absorption frequency characteristic quite unlike that of a new piece.

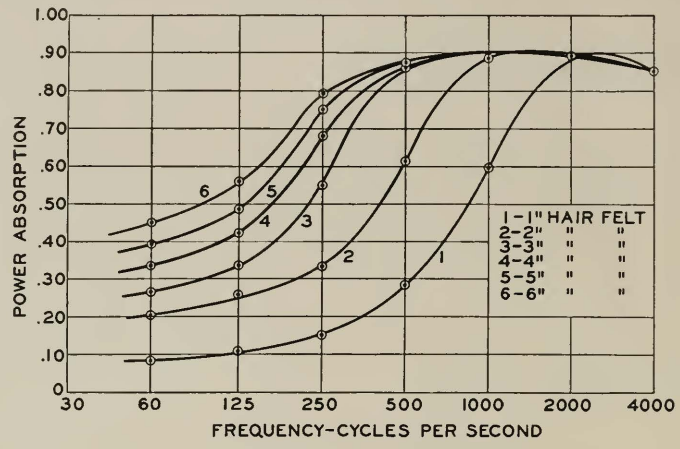


Fig. 4—Power absorbed by hair felt

In Fig. 4 are given the absorption coefficients for various thicknesses of hair felt. These values are in general agreement with those obtained by the reverberation method according to published results. Exact agreement is not to be expected, for the values here given apply only to sound waves having a perpendicular incidence on materials solidly backed by a hard surface. When the materials are applied in a room, the support is often more flexible and the absorption is partly due to inelastic bending. However, the agreement between the sets of values is sufficiently good to show that the results obtained by the simpler tube method may be used to get a good approximation to the values of absorption of the materials when applied in rooms for damping purposes.

Measurements have been made on a large number of porous materials. Although most of these materials are very good absorbers at the higher frequencies, none of them were found to be very efficient in the lower frequency region. Uniform absorption over most of the frequency range was found only in materials which are relatively inefficient absorbers. High absorption at the lower frequencies was obtained only when the thickness of the material was greatly increased. This fact is typically illustrated by the curves of absorption for hair felt given in Fig. 4.

When a sound wave of low frequency is reflected from a wall covered

TABLE I

ABSORPTION COEFFICIENTS FOR VARIOUS FREQUENCIES

FREQUENCY C.P.S.

	60	125	250	500	1000	2000	4000
1″ Acoustic tile	.07	.08	.11	.18	.48	.76	.47
½″ Asbestos hair felt	.08	.10	.15	.24	.49	.84	.66
Axminster rug	.07	.11	.14	.20	.33	.52	.82
Felted wood fibre	.08	.12	.18	.33	.67	.92	.91
½″ Building board	.13	.14	.15	.17	.20	.26	.29
Flax wool	.07	.09	.18	.48	.73	.50	.33
Structure No. 1	.12	.18	.36	.71	.79	.82	.85
" " 2	.16	.24	.46	.77	.92	.89	.85
" " 3	.23	.37	.62	.88	.91	.78	.84
" " 4	.19	.28	.51	.81	.92	.90	.84
" " 5	.15	.25	.44	.75	.77	.71	.80
" " 6	.22	.41	.87	.74	.81	.59	.83
" " 7	.17	.39	.82	.94	.92	.91	.85
" " 8	.24	.39	.83	.82	.64	.59	.80
" " 9	.22	.37	.79	.91	.82	.89	.86
" " 10	.30	.55	.92	.69	.83	.86	.86
Structure No. 1	Fiber building board —no air space—felt						
" " 2	Fiber building board —1″ air space—felt						
" " 3	Fiber building board —2″ air space—felt						
" " 4	Fiber building board —no air space—2″ felt						
" " 5	Fiber building board —1″ air space—Fiber building board						
" " 6	Fiber building board —1″ air space— Fiber building board—1″ air space—Fiber building board						
" " 7	Fiber building board —1″ air space—felt—1″ air space—felt						
" " 8	Fiber building board —1″ air space—felt—1″ air space—Fiber building board						
" " 9	Fiber building board —1″ air space— Fiber building board—1″ air space—felt						
" " 10	Fiber building board —1″ air space— Fiber building board—1″ air space—Fiber building board—felt						

with a porous material, the velocity of the air particles near the reflecting surface is small and hence there can be but little absorption. We may look at the phenomenon of reflection in still another way. In order to have a small coefficient of reflection the mechanical impedance of the wall per unit area should, as nearly as possible, be equal to the acoustic impedance of the air per unit area. The reason for the high reflection at low frequencies by a rigid wall covered with a porous material lies in its high stiffness reactance. At a given frequency this reactance can be compensated by loading the air near the reflecting surface. This may be accomplished in various ways. One of these ways is to place at a short distance from the wall a second wall which is porous or perforated. This arrangement has the effect of covering the wall with a multiplicity of resonators, which may be given any desired resonance frequency by properly proportioning the size, length and number of perforations and the spacing of the walls. The surface of the walls forming the air space should be absorbing or else the space should be provided with absorbing material.

To get a wider absorption band two or more perforated walls with proper spacing may be used, as this arrangement is equivalent to an aggregate of multiple resonators. The values of absorption coefficients of a number of structures of this type are given in the accompanying table. The measurements refer to sound which is incident from right to left as the structures are given in the table. The building board referred to in the table is a commercial type of insulating-board one inch thick with 400 1/4 inch by 3/4 inch holes per square foot. The felt in all cases is one-inch hair felt. These values show that relatively high absorption may be obtained at low as well as at high frequencies without an excessive amount of absorbing material. The use of combinations of absorbing materials, such as are given in the table, offers the advantage that more uniform damping at all frequencies can be obtained, and the degree of damping can be readily controlled by covering the proper area of surface. These two factors have become increasingly important in studio and auditorium design, with improved technique in recording and reproducing speech and music.

The Rigorous and Approximate Theories of Electrical Transmission Along Wires

By JOHN R. CARSON

THE theory of electrical transmission along straight parallel guiding conductors is of fundamental importance to the communication engineer. In its original, and largely in its present day form, it involves only relatively simple concepts which go back to the early work of Kelvin and Heaviside. In accordance with these concepts the transmission phenomena are completely determined by the self and mutual impedances of the conductors and the self and mutual capacities (together with the dielectric leakage). As a consequence, the phenomena are completely expressed in terms of the propagation constants and corresponding characteristic impedances of the possible modes of propagation deducible from these underlying concepts.

The elementary theory sketched above is of beautiful simplicity and great value. It is, however, admittedly approximate, and in two respects is not altogether adequate. Its first defect is that it represents the transmission phenomena correctly only at some distance from the physical terminals of the system or at some distance from points of discontinuity. This defect is ordinarily of small practical significance when the conductors all consist of wires of small cross section. When, however, conductors of large cross sections, or the ground, form part of the transmission system, the elementary theory may be quite inadequate. The theoretical questions here involved were briefly discussed by the writer in a previous paper.[1] The mathematics involved in this problem are extremely complicated and the further work of the writer has not as yet been carried to a point which justifies publication.

With the extension of transmission theory discussed in the preceding paragraph the present paper has no concern, and it is to be expressly understood that we are dealing with the transmission phenomena at a sufficient distance from the physical terminals, such that the "end effects" are negligible. The problems here dealt with may be stated as follows: First to investigate the conditions under which the specification of the system by means of its self and mutual impedances is valid and secondly to provide a general method for calculating these circuit parameters from the geometry and electrical constants of the system.

[1] "The Guided and Radiated Energy in Wire Transmission." Trans. A. I. E. E., 1924.

As regards the first phase of this problem it is found that the complete specification of the system in terms of its self and mutual impedances and capacities is only rigorously valid for the ideal case of perfect conductors embedded in a perfect dielectric, and that it becomes quite invalid if either the conductors or the dielectric are too imperfect. Fortunately, however, it is valid to a high degree of approximation for all systems which could be employed for the *efficient* transmission of electrical energy.

Under the circumstances where the approximations discussed in the preceding paragraph are valid it is shown that the electric and magnetic field in both dielectric and conductors are derivable from two wave functions. The first of these is determined as a linear function of the conductor charges by the solution of a well-known two-dimensional potential problem, while the second is determined as a linear function of the conductor currents by the solution of a generalization of the two-dimensional potential problem. The latter problem is believed to be novel, in its general form, and to possess both practical and mathematical interest. For detailed application of the theory to specific problems, the following papers may be consulted.

"Wave Propagation over Parallel Wires: The Proximity Effect." *Phil. Mag.*, April 1921.

"Transmission Characteristics of the Submarine Cable." *Jour. Frank. Inst.*, Dec. 1921.

"Wave Propagation in Overhead Wires with Ground Return." *B. S. T. J.*, Oct. 1926.

I

Maxwell's equations are the set of partial differential equations which formulate the relations between the electric intensity E and the magnetic intensity H in terms of the frequency $\omega/2\pi$ and the electrical constants of the medium. Let λ, μ and k denote the conductivity, permeability and dielectric constant of the medium; let it be supposed that all quantities vary with the time t as $e^{i\omega t}$, and let

$$v = 1/\sqrt{k\mu},$$
$$\nu^2 = 4\pi\lambda\mu i\omega - \omega^2/v^2,$$
$$i = \sqrt{-1}.$$

Then if we introduce the vector

$$M = \mu i\omega \cdot H,$$

Maxwell's equations for a *continuous homogeneous* medium may be written in the compact form [2]

$$\begin{aligned} \text{curl } E &= -M, \\ \text{curl } M &= \nu^2 E, \\ \text{div } E &= 0, \\ \text{div } M &= 0. \end{aligned} \tag{1}$$

From this set of equations it is easily shown that each component of the vectors E and M individually satisfies the *wave equation*

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \nu^2\right) f = 0 \tag{2}$$

or in vector notation

$$(\nabla^2 - \nu^2) f = 0.$$

Here f denotes any vector component; thus in Cartesian coordinates f may stand for E_x, E_y, E_z; M_x, M_y, M_z, all of which separately satisfy (2).

Given the electrical constants and geometry of the conducting system and dielectric media, the general problem is to find solutions of (1) and (2) which also satisfy the *boundary conditions* at the surfaces of separation of the different media. These boundary conditions are that the tangential components of E and H shall be continuous over such surfaces of separation. These boundary conditions, as may be seen from (1), necessitate also the continuity of the *normal* components of M and $(\nu^2/\mu)E$.

If we introduce a vector potential $A(A_{x,\,y,\,z})$ and a scalar potential Φ, it is easily shown that (1) may be replaced by

$$\begin{aligned} M &= \text{curl } A, \\ E &= -A - \text{grad } \Phi, \end{aligned} \tag{3}$$

with the further relation

$$\text{div } A + \nu^2 \Phi = 0. \tag{4}$$

Φ and the components of the vector A individually satisfy the wave equation; thus

$$\begin{aligned} (\nabla^2 - \nu^2)\Phi &= 0, \\ (\nabla^2 - \nu^2) A &= 0. \end{aligned} \tag{5}$$

In Cartesian coordinates these equations are

[2] Note that in this form the constants of the medium appear explicitly only through the parameter ν^2.

$$
\begin{aligned}
M_x &= \frac{\partial}{\partial y} A_z - \frac{\partial}{\partial z} A_y, \\
M_y &= \frac{\partial}{\partial z} A_x - \frac{\partial}{\partial x} A_z, \\
M_z &= \frac{\partial}{\partial x} A_y - \frac{\partial}{\partial y} A_x, \\
E_x &= -A_x - \frac{\partial}{\partial x} \Phi, \\
E_y &= -A_y - \frac{\partial}{\partial y} \Phi, \\
E_z &= -A_z - \frac{\partial}{\partial z} \Phi,
\end{aligned}
\tag{6}
$$

and

$$
\frac{\partial}{\partial x} A_x + \frac{\partial}{\partial y} A_y + \frac{\partial}{\partial z} A_z + \nu^2 \Phi = 0. \tag{7}
$$

In technical transmission problems we are largely concerned with propagation along a uniform transmission system, composed of straight parallel conductors. That is to say, the transmission system does not vary geometrically or in its electrical constants along the axis of transmission, taken as the axis of Z. It is known that in such transmission systems *exponentially* [3] propagated waves exist. We therefore modify the general equations by assuming that the wave (and all vector components) vary with t and z as exp $(i\omega t - \gamma z)$, γ being entitled the *propagation constant.* As a consequence of this assumption it is easily shown that the vectors E and M are derivable from the wave functions F, Φ, Θ as follows:

$$
\begin{aligned}
M_x &= \frac{\partial}{\partial y} F - \gamma \frac{\partial}{\partial x} \Theta, \\
M_y &= -\frac{\partial}{\partial x} F - \gamma \frac{\partial}{\partial y} \Theta, \\
M_z &= -(\nu^2 - \gamma^2)\Theta, \\
E_x &= -\frac{\partial}{\partial x} \Phi - \frac{\partial}{\partial y} \Theta, \\
E_y &= -\frac{\partial}{\partial y} \Phi + \frac{\partial}{\partial x} \Theta, \\
E_z &= -\frac{\partial}{\partial z} \Phi - F = \gamma \Phi - F.
\end{aligned}
\tag{8}
$$

The *wave functions* F and Φ are not independent but are connected by the relation

$$
\nu^2 \Phi = \gamma F. \tag{9}
$$

[3] This means that the wave involves the axial coordinate z only exponentially.

Another useful formulation of the field equations equivalent to and directly deducible from (8) is

$$\begin{aligned}
(\nu^2 - \gamma^2)M_x &= -\nu^2 \frac{\partial}{\partial y} E_z + \gamma \frac{\partial}{\partial x} M_z, \\
(\nu^2 - \gamma^2)M_y &= \nu^2 \frac{\partial}{\partial x} E_z + \gamma \frac{\partial}{\partial y} M_z, \\
(\nu^2 - \gamma^2)E_x &= \gamma \frac{\partial}{\partial x} E_z + \frac{\partial}{\partial y} M_z, \\
(\nu^2 - \gamma^2)E_y &= \gamma \frac{\partial}{\partial y} E_z - \frac{\partial}{\partial x} M_z.
\end{aligned} \tag{10}$$

In this formulation the problem is reduced to the determination of the *wave functions* E_z and M_z, and the *propagation constant* γ.

It will be observed that, by virtue of the assumption that the wave functions of (8), (9) and (10) involve t and z only through the common factor $\exp(i\omega t - \gamma z)$, we can write

$$\begin{aligned}
F &= f(x, y) \cdot \exp(i\omega t - \gamma z), \\
\Phi &= \phi(x, y) \cdot \exp(i\omega t - \gamma z), \\
E &= e(x, y) \cdot \exp(i\omega t - \gamma z), \text{ etc.},
\end{aligned} \tag{11}$$

where f, ϕ, e, etc., are two-dimensional functions of x and y alone, and satisfy the two-dimensional wave equations

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) f = (\nu^2 - \gamma^2) f, \text{ etc.} \tag{12}$$

In the following, therefore, we shall regard the wave functions F, Φ, E, etc., as two-dimensional functions with the understanding that the common factor $\exp(i\omega t - \gamma z)$ is omitted for convenience.

II

Before taking up the discussion of the general problem in the light of equations (8) and (10) we shall first consider a type of plane wave propagation to which the transmission phenomena closely approximate in an efficient transmission system. We consider the ideal transmission system composed of any number of straight parallel *perfectly conducting* conductors imbedded in a *perfect* dielectric. For such a system we *assume* the possibility of plane wave propagation by supposing that E_z and M_z are everywhere zero. By virtue of the assumption of perfect conductivity, the electric force must vanish inside the conductors, and at the surface the tangential component

must vanish. *In the dielectric* reference to equations (10) shows that if $E_z = M_z = 0$, a finite solution requires that

$$\nu^2 - \gamma^2 = 0$$

or, since $\lambda = 0$ in the dielectric,

$$\gamma = i\omega/v.$$

That is to say, the plane wave is propagated with the velocity of light v, without attenuation.

Reference to equations (8) and (9) shows that the boundary conditions can be satisfied by setting $\Theta = 0$, writing

$$\Phi = \phi \cdot \exp(i\omega t - i\omega z/v),$$

and determining the function ϕ which satisfies Laplace's equation in two dimensions,

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\phi = 0,$$

and is constant over the cross section of the conductors.

From the relation $F = (i\omega/v)\Phi$ it is also easily shown that the electric and magnetic forces are both in planes normal to Z and that these vectors are normal to each other and in time phase. The flow of energy is therefore parallel to the Z-axis everywhere. We therefore have a pure plane guided wave of unit power factor; the ideal for the electrical transmission of energy.

III

We now take up the much more complicated problem arising when the conductivity λ of the conductors is finite and when the dielectric media themselves may be dissipative. In attacking this general problem we shall be guided throughout by the fact that the wave solution we are seeking must approximate, more or less closely, to the ideal plane wave [4] if the system is to efficiently transmit electrical energy. We shall therefore introduce *ab initio* approximations which must be valid in all efficient transmission systems. These approximations cannot be all justified *a priori;* their justification must come *a posteriori* from the fact that the final solution satisfies the original assumptions and approximations.

[4] It is to be noted that the solution sought is the *principal wave.* (See "The Radiated and Guided Energy in Wire Transmission," *Trans. A. I. E. E.*, 1924.) This wave does not, in general, completely represent the phenomena, except at a considerable distance from the physical terminals of the transmission system, and then only in the neighborhood of the conductors.

First we have to define what we mean by conductor and by dielectric; the significance of these definitions will appear in the course of the analysis. *A conducting medium* is one in which ω^2/v^2 is very small compared with $4\pi\lambda\mu\omega$; while a *dielectric medium* is one in which $4\pi\lambda\mu\omega$ is very small compared with ω^2/v^2. The intermediate cases will not be discussed in the present paper; in the following it will be assumed that the conductors and dielectrics satisfy these definitions.[5]

The assumptions which we make at the outset in the approximate solution may now be listed and qualitatively justified as follows:

1. The propagation constant γ is an extremely small quantity and its real part is not large compared with its imaginary part. Since $|\gamma|$ is of the order of magnitude of $\omega \cdot 10^{-10}$, it is evident that γ is very small even for frequencies of millions of cycles per second. As regards the second restriction, if the real part of γ is large compared with the imaginary, the wave will be damped out in a few wave-lengths, and the system cannot efficiently transmit energy.

2. *In the conductors* the axial electric intensity E_z is large compared with the component normal to Z. This restriction means that the dissipation in the conductors due to the axial currents is large compared with the dissipation due to the charging currents. Evidently this restriction is necessary for the efficient transmission of energy.

3. *In the dielectric* the axial electric intensity is small compared with the normal electric intensity. The justification of this assumption is as follows: The propagation of energy occurs in the dielectric, and is normal to the direction of the electric intensity. Since the usefully transmitted energy is propagated along the axis of transmission and the propagation normal to the axis simply means dissipation, the axial electric intensity must be small compared with the normal component for efficient transmission.

4. The axial magnetic intensity H_z is everywhere small compared with the normal intensity. The justification of this assumption depends on the same arguments as (3).

As regards (3) and (4) it will be remarked that in the ideal plane wave propagation both E_z and M_z are zero. In the case of imperfect conductors E_z in the dielectric is not zero but may be regarded as a first order small quantity. M_z on the other hand is to be regarded as a second order small quantity because it not only vanishes for the case of perfect conductors but also vanishes for the case of imperfect conductors for the case where the wave is made up of a set of compo-

[5] In accordance with these definitions, conductors and dielectrics depend for their classifications on the frequency, as well as their electrical constants. The definition of *conductor* means that the displacement current is negligible compared with the conduction current.

nent radially symmetrical waves oriented on the axes of the conductors; to this the actual wave approximates in important transmission systems.

We shall now introduce the consequences of the foregoing assumptions into the differential equations of the problem.

IV

Referring to equations (10), these may be replaced *in the conductors only* where γ^2 is very small compared with ν^2 and γ is a very small quantity, by the approximation:

$$\begin{aligned} M_x &= -\frac{\partial}{\partial y} E_z, \\ M_y &= \frac{\partial}{\partial x} E_z, \\ E_x &= \frac{\gamma}{\nu^2}\left\{\frac{\partial}{\partial x} E_z + \frac{1}{\gamma}\frac{\partial}{\partial y} M_z\right\}, \\ E_y &= \frac{\gamma}{\nu^2}\left\{\frac{\partial}{\partial y} E_z - \frac{1}{\gamma}\frac{\partial}{\partial x} M_z\right\}, \end{aligned} \tag{13}$$

Therefore *in the conductors* the vector components M_x, M_y are derivable by spatial differentiation from E_z. E_x, E_y are not in general so derivable on account of the factor $1/\gamma$, a very large quantity, which appears with M_z. (It appears that γE_z and M_z may be of comparable orders of magnitude.) We assume, however, for reasons discussed above, that both E_x and E_y are very small compared with E_z *in the conductors.*

In the dielectric, where ν^2 and γ^2 are of comparable orders of magnitude, the foregoing approximations are not valid and the rigorous equations must be employed. Returning to equations (10) and writing for convenience $\gamma^2/\nu^2 = \beta$, we have

$$\begin{aligned} M_x &= -\frac{1}{1-\beta}\left\{\frac{\partial}{\partial y} E_z - \frac{\beta}{\gamma}\frac{\partial}{\partial x} M_z\right\}, \\ M_y &= \frac{1}{1-\beta}\left\{\frac{\partial}{\partial x} E_z + \frac{\beta}{\gamma}\frac{\partial}{\partial y} M_z\right\}, \\ E_x &= \frac{\beta}{\gamma}\frac{1}{1-\beta}\left\{\frac{\partial}{\partial x} E_z + \frac{1}{\gamma}\frac{\partial}{\partial y} M_z\right\}, \\ E_y &= \frac{\beta}{\gamma}\frac{1}{1-\beta}\left\{\frac{\partial}{\partial y} E_z - \frac{1}{\gamma}\frac{\partial}{\partial x} M_z\right\}. \end{aligned} \tag{14}$$

In equations (13) and (14), x and y may be any orthogonal coordinate system. Let us suppose that they are so chosen that x is

tangential to the conductor surface; M_y is therefore the normal component of M at the surface of the conductor and must there be continuous. E_z and $(\partial/\partial x)E_z$ are also continuous. Consequently, we must have by equating M_y as given by (13) and (14),

$$\frac{1}{\gamma}\left(\frac{\partial}{\partial y}M_z\right)_e = -\frac{\partial}{\partial x}E_z, \tag{15}$$

the subscript e indicating the value of $(\partial/\partial y)M_z$ outside the conductor. But from the expression for E_x, as given by (14), this is precisely the condition that makes $E_x = 0$ at the surface of the conductor. Consequently we arrive at the very important proposition that, subject to the approximations involved in (13), *the tangential component of E in the xy-plane vanishes at the conductor surfaces.*

We shall now find it convenient to express the field in the dielectric in accordance with (8) in terms of the wave functions F, Φ, Θ. Writing

$$\Theta = \theta \cdot \exp(i\omega t - \gamma z), \tag{16}$$

θ satisfies the differential equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\theta = (\nu^2 - \gamma^2)\theta. \tag{17}$$

Now, *in the dielectric*, ν^2 and γ^2 are both exceedingly small quantities which are nearly equal, so that $\nu^2 - \gamma^2$ is the difference of two very small and nearly equal quantities. We therefore replace it by zero, so that

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\theta = 0. \tag{18}$$

θ is therefore a two-dimensional potential function. Consequently *a conjugate two-dimensional potential function ψ exists, such that*

$$\begin{aligned} \frac{\partial}{\partial x}\theta &= \frac{\partial}{\partial y}\psi, \\ \frac{\partial}{\partial y}\theta &= -\frac{\partial}{\partial x}\psi. \end{aligned} \tag{19}$$

Writing

$$\Psi = \psi \cdot \exp(i\omega t - \gamma z),$$

equations (8) become

$$M_x = \frac{\partial}{\partial y}(F - \gamma\Psi),$$

$$\begin{aligned} M_y &= -\frac{\partial}{\partial x}(F - \gamma\Psi), \\ E_x &= -\frac{\partial}{\partial x}(\Phi - \Psi), \\ E_y &= -\frac{\partial}{\partial y}(\Phi - \Psi), \\ E_z &= \gamma(\Phi - \Psi) - (F - \gamma\Psi). \end{aligned} \tag{20}$$

Introducing new wave functions

$$\begin{aligned} F' &= F - \gamma\Psi, \\ \Phi' &= \Phi - \Psi, \end{aligned} \tag{21}$$

we have (dropping primes)

$$\begin{aligned} M_x &= \frac{\partial}{\partial y}F, \\ M_y &= -\frac{\partial}{\partial x}F, \\ E_x &= -\frac{\partial}{\partial x}\Phi, \\ E_y &= -\frac{\partial}{\partial y}\Phi, \\ E_z &= \gamma\Phi - F, \end{aligned} \tag{22}$$

where now Φ and F are *independent* wave functions.

If the foregoing analysis has been carefully followed, the important advantage of equations (22) as compared with (8) will be appreciated. The transformation of (8) into (22) is strictly dependent upon and conditioned by the legitimacy of neglecting $\nu^2 - \gamma^2$ in the dielectric, whereby the wave functions are essentially reduced to two-dimensional potential functions. It is evident that the whole engineering theory of transmission involves this approximation.

V

We are now prepared to sketch the general solution of the problem,[6] employing equations (13) *in the conductors*, and equations (22) *in the dielectric*. The procedure is as follows:

1. At the surfaces of the conductors the tangential component E_τ in the xy-plane of E vanishes, as shown above. That is,

$$E_\tau = -\frac{\partial}{\partial \tau}\Phi = 0 \tag{23}$$

[6] For detailed applications of this method of solution to specific problems, the published papers referred to in the introduction to this paper may be consulted.

at the surface of each conductor.[7] In the dielectric outside the conductor, the potential Φ satisfies Laplace's equation in two dimensions; hence

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\Phi = 0 \tag{24}$$

in the dielectric; and

$$\frac{\partial}{\partial \tau_j}\Phi = 0, \qquad (j = 1, 2, \cdots n) \tag{25}$$

at the surface of the jth conductor. Also

$$\oint E_n{}^j d\tau_j = -\int \frac{\partial}{\partial n_j}\Phi d\tau_j = \frac{4\pi}{k} Q_j, \qquad (j = 1, 2, \cdots n) \tag{26}$$

the integration being carried around the surface of the jth conductor, Q_j being the charge per unit length on the jth conductor.

The determination of Φ from (24)–(26), when the geometry of the conductors is specified, is a well-known two-dimensional potential problem, for the solution of which very general methods are available. The solution results in the form

$$\Phi = \phi_1(x, y)Q_1 + \phi_2(x, y)Q_2 + \cdots + \phi_n(x, y)Q_n. \tag{27}$$

That is, Φ is a linear function of the conductor charges $Q_1 \cdots Q_n$, and the coefficients $\phi_1 \cdots \phi_n$ are unique functions of the geometry of the transmission system and are determinable by the usual methods of two-dimensional potential theory.

2. The continuity of M_n and $(1/\mu)M_\tau$ at the surfaces of the conductors is analytically formulated by the equations

$$\begin{aligned} \frac{\partial}{\partial \tau} F &= -\frac{\partial}{\partial \tau} E_z, \\ \frac{\partial}{\partial n} F &= -\frac{\mu}{\mu_c}\frac{\partial}{\partial n} E_z, \end{aligned} \tag{28}$$

where μ is the permeability of the dielectric and μ_c that of the conductor. These relations, it will be understood, hold at the surfaces of all the conductors. F is a wave function which satisfies Laplace's equation in two dimensions *in the dielectric;* thus

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) F = 0, \tag{29}$$

[7] In the following, τ and n denote vectors tangential and normal to the conductor surface respectively.

and E_z is a wave function which *in the conductor* satisfies the two-dimensional wave equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) E_z = \nu^2 E_z. \tag{30}$$

In addition, E_z and F are connected with the conductor current I by the relations

$$\begin{aligned} I &= \lambda \int E_z dS, \\ 4\pi\mu i\omega \cdot I &= \oint \frac{\partial}{\partial n} F d\tau, \end{aligned} \tag{31}$$

λ here denoting the conductivity of the conductor.

It follows at once that the determination of F and E_z from (28)–(31) is a generalization of the two-dimensional potential problem involved in the determination of Φ from (24)–(26); it may be precisely stated as follows:

The function F satisfies Laplace's equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) F = 0 \tag{32}$$

everywhere outside the n conductors. Inside the jth conductor the electric force E_z^j satisfies the two-dimensional wave equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) E_z^j = \nu_j^2 E_z^j, \qquad (j = 1, 2, \cdots n) \tag{33}$$

while at the surface of the jth conductor

$$\begin{aligned} \frac{\partial}{\partial \tau_j} F &= -\frac{\partial}{\partial \tau_j} E_z^j, \\ \frac{\partial}{\partial n_j} F &= -\frac{\mu}{\mu_j} \frac{\partial}{\partial n_j} E_z^j, \qquad (j = 1, 2, \cdots n) \end{aligned} \tag{34}$$

and

$$4\pi\mu i\omega \cdot I_j = \oint \frac{\partial}{\partial n_j} F d\tau_j, \qquad (j = 1, 2, \cdots n). \tag{35}$$

Just as equations (24)–(26) uniquely determine Φ as a linear function of $Q_1 \cdots Q_n$, so equations (32)–(35) uniquely determine the potential function F in the dielectric and the electric intensities $E_z^{(1)} \cdots E_z^{(n)}$ in the n conductors as linear functions of the conductor currents; thus

$$F = f_1(x, y) I_1 + f_2(x, y) I_2 + \cdots + f_n(x, y) I_n, \tag{36}$$

and

$$E_z^j = e_{1j}(x, y)I_1 + \cdots + e_{nj}(x, y)I_n, \qquad (j = 1, 2, \cdots n) \quad (37)$$

the f and e functions depending on the geometry of the conducting system and, through the parameter ν^2, on its electrical constants. They are uniquely determined by the differential equations.

The actual solution of the differential equations (32)–(35) is essentially more difficult than the solution of (24)–(26) involved in the determination of Φ, and they have not been subjected to the exhaustive study accorded to the potential problem. On the other hand, the analogy with the potential problem suggests that extension and modifications of the general and well-known methods of solution available for that problem should be possible.

To summarize the foregoing we have succeeded in expressing the potential function Φ (and therefore E_x, E_y) *in the dielectric* as a linear function of the conductor charges, the coefficients of the conductor charges $Q_1 \cdots Q_n$ being spatial functions of x and y which we determined by the usual methods of two-dimensional potential theory.

Similarly it has been shown that E_z (and therefore the current distribution) *in the conductors*, and the potential function F (and therefore the magnetic field) *in the dielectric*, are expressible as linear functions of the conductor currents $I_1 \cdots I_n$, the determination of the coefficients depending on the solution of a generalization of the two-dimensional potential problem.

3. To complete the solution of the problem, recourse is had to the fact that E_z is continuous at the surfaces of the conductors. At the surface of each conductor we therefore equate E_z, as given by (37), in terms of the currents $I_1 \cdots I_n$ with $\gamma\Phi - F$ (see (22)), Φ being given by (27) in terms of $Q_1 \cdots Q_n$ and F by (36) in terms of $I_1 \cdots I_n$. This gives n equations of the form

$$\begin{aligned} Z_{11}I_1 + \cdots + Z_{n1}I_n &= \gamma\Phi_1 = \gamma(p_{11}Q_1 + \cdots + p_{1n}Q_n), \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ Z_{n1}I_1 + \cdots + Z_{nn}I_n &= \gamma\Phi_n = \gamma(p_{n1}Q_1 + \cdots + p_{nn}Q_n). \end{aligned} \quad (38)$$

Here $\Phi_1 \cdots \Phi_n$ are the values of Φ at the surfaces of the n conductors respectively; the p coefficients are Maxwell's potential coefficients of the system, while the Z coefficients are the self and mutual "impedances" of the conductors.

We require a further relation between I and Q; this is furnished by the well-known relation

$$\begin{aligned} i\omega Q &= \gamma I - \lambda \oint E_n ds \\ &= \gamma I + \lambda \oint \frac{\partial}{\partial n} \Phi ds \\ &= \gamma I - \frac{4\pi\lambda}{k} Q, \end{aligned} \tag{39}$$

the integration being carried around the contour of the conductor. (λ is the conductivity of the dielectric and the last term is the "leakage" current.) We have therefore, for a homogeneous dielectric,

$$\left(i\omega + \frac{4\pi\lambda}{k} \right) Q = \gamma I, \tag{40}$$

which furnishes the necessary relation.

Elimination of Q from (38) by means of (40) gives n homogeneous equations in $I_1 \cdots I_n$, the coefficients involving only one unknown quantity, the propagation constant γ. A finite solution necessitates the vanishing of the determinant of the coefficients; equating this to zero gives an nth order equation in γ^2, which determines the n possible values of γ, and therefore the n possible modes of propagation in the system. The formal solution of the problem is thus completed.

In conclusion it is worth while reviewing and summarizing the mathematical restrictions on the solution developed in the foregoing pages; restrictions which have their counterpart in the physical requirements of the system for the efficient guided transmission of electromagnetic energy. The essential restrictions are that (1) *in the conductors* γ^2 is very small compared with ν^2, and (2) *in the dielectric* the wave equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \Phi = (\nu^2 - \gamma^2)\Phi$$

may be replaced, at least in the neighborhood of the conductors, by

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \Phi = 0.$$

If the conductors are so imperfect, or the dielectric so dissipative that these approximations are not justified, the method of solution

given above breaks down, and the problem must be attacked from the rigorous equations. These have never been solved in general, in fact the only rigorous solution known to the writer is for the case of circular symmetry and even this involves the location of the roots of an extremely complicated transcendental equation. Fortunately, in view of these difficulties, the general case of quite imperfect conductors or imperfect dielectric media is of small technical importance for the reason given above.

Some General Results of Elementary Sampling Theory for Engineering Use

By PAUL P. COGGINS

EVERY day we base conclusions on the results of the process commonly known as "sampling." For example, if five times in a week a man has waited ten minutes or more for his trolley at a street corner, he may conclude that the transportation facilities are poor. Or again, if a housewife has bought ten loaves of bread at a certain store and has found five of them not as fresh as might be desired, she decides that in the future she will buy her bread elsewhere. Both of these conclusions are based on an intuitive application of sampling theory. Such examples could be multiplied indefinitely.

Similarly, in most engineering problems, observational data are involved in one way or another. In order to be able to assign the proper significance to these data, it is essential to have some idea as to their reliability, that is, to what extent they represent all the facts under consideration. First, the measurements themselves may be in error. In the second place, although the observations may have been made with perfect precision, they may be incomplete; they may constitute but a "sample" of a large group of possible observations. The problem considered in this paper is one of this second class, generally known as "sampling" problems.

Assume the existence of a total group or "universe" of N objects and that observations have been made on a certain number n of them with reference to a particular characteristic. This number n we will call the "sample." From this sample we wish to deduce some estimate concerning the probable condition of that universe with reference to the characteristic observed.

Now the characteristic observed may itself take on one of two forms. It may be either, (1) present or absent; (2) quantitative. For simplicity in discussion we may call the first, "Sampling of Attributes," and the second, "Sampling of Variables."

An example of each will be cited from the telephone field.

Example 1: Sampling of Attributes

Suppose that 4,000 relays of a particular type constitute a day's output. In order to determine roughly what proportion of these are non-operative at a current of 12 mils, a sample of 500 relays is tested and out of this sample 10 fail to operate at the required current. In

the sample, then, two per cent of the relays were defective. What, then, is the probability that the percentage of the 4,000 relays having this defect is between one and three per cent? Or what is the probability that the percentage of defectives in the universe of 4,000 does not exceed four per cent? Or again, if we wish to be practically certain that among the 4,000 relays not more than two per cent are defective in this respect, how many defectives would be allowable in a sample of 200? or a sample of 1,000? Any number of questions of this sort can be asked and may be answered on the basis of the proper assumptions by sampling theory.

Example 2: Sampling of Variables

An office serves 5,000 subscribers lines. Measurements of the insulation resistance are made on 200 of these, selected at random, and the resulting values tabulated. They vary all the way from 12,000 ohms to 200,000 ohms. What conclusions may be drawn as to the probability that more than a certain number, say 20 of the subscribers' loops out of the 5,000, have an insulation resistance of less than 18,000 ohms? What is the most probable distribution of the insulation resistances for the office as a whole? What is the probable error of the average of the observations as a measure of the average loop insulation resistance for the office?

As before, much information *regarding the universe* may be inferred from a properly chosen sample, always, however, with some degree of uncertainty. This uncertainty, so far as the sampling process is concerned, naturally decreases as the size of the sample increases, and, of course, disappears except for inaccuracies of measurement, when the sample becomes coextensive with the universe.

The respective treatments of these two types of problems differ considerably in detail. The basic principles are, however, essentially the same, and involve in each case the notions of "a posteriori" probability, as discussed in most of the standard textbooks on the theory of probability.

In both problems there are certain observations. By means of these we desire to obtain as precise information as possible concerning some one or more characteristics of the universe from which these observations or samples were drawn. The true nature of the universe is to some degree, at least, unknown. Certain hypotheses concerning it may, however, in the light of the sample be more probable than others. What we wish to estimate is the probability that either a particular hypothesis or a group of mutually exclusive hypotheses includes the true one.

This article will be devoted to the type of problem termed "Sampling of Attributes." [1] In it are included results from an extensive series of computations in the form of charts which may be of value in the solution of practical engineering problems. The nomenclature is general, so as to be applicable to a wide variety of practical problems. For convenience in discussion we shall divide the units of any sample into the two mutually exclusive classes, "defective" and "satisfactory." The following notation will be used:

N = total number of items in universe,

n = total number of items in sample,

X = number of defective items in universe (unknown),

c = number of defective items in sample (observed),

$w(X)$ = *a priori* probability that the universe will contain exactly X defectives,

$W(X_1, X_2)$ = *a posteriori* probability that the universe contains a number of defectives X such that $X_1 \leqq X \leqq X_2$.

It is of extreme importance that, at the outset, the significance of the symbol $w(X)$ in sampling problems be clearly defined. It is a measure of the probability, *before the sample* is taken, that the lot or universe in question contains X defective items and $N - X$ satisfactory items. It may be based on previous samples, or the reputation of the manufacturer producing those items, or on any one or more of a number of other pertinent data. For example, even before a sampling inspection, we should unhesitatingly say that in a lot of 1,000 relays sent out by a reputable manufacturer it is very much more likely, *a priori*, that the lot will contain less than 100 relays with a short-circuited winding than that the lot will contain more than 800 relays defective in the same respect. We should probably find ourselves in a quandary, however, if we attempted to state without a sample inspection, the relative likelihoods of 3, 4, 5, 6, $\cdots$, etc., defectives existing in the lot. $w(X)$ is a function whose numerical value is assumed to state this *a priori* probability. The extent to which we are able to make use of this function, then, depends on how precisely we are able to assign numerical values to it before we study our sample.

[1] This general type of problem has been under study within the Bell System for some time. In an article "Deviation of Random Samples from Average Conditions and Significance to Traffic Men" by E. C. Molina and R. P. Crowell which appeared in the *Bell System Technical Journal* for January, 1924, a special case of sampling theory was developed and various possible applications were suggested. In August, 1924, Molina delivered a paper entitled "A Formula for the Solution of Some Problems in Sampling" before the statistical section of the International Mathematical Congress in Toronto, Canada. This paper dealt with a somewhat more general case of the sampling problem than was discussed in the article just mentioned.

It may be helpful at this point to state and solve a simple problem, which will serve to bring out the fundamental principles involved. An urn is known to contain 10 balls, some of which are white and the others black. Five balls have been drawn and *not* replaced. Of these five, one is white and four are black. What is now the probability that the urn originally contained just one white ball and nine black? Two white and eight black?

Before we proceed to obtain a solution for this problem we have to make some assumption, based on knowledge available before the drawings were made, concerning the probability that the urn contains black and white balls in any given proportion.

Consider two such assumptions—

(*a*) All proportions are *a priori* equally likely, i.e., before the drawings it is as likely that three whites and seven blacks were put in the urn as six whites and four blacks, etc.

(*b*) The urn was filled with ten balls drawn at random from a bag containing a very large number of balls of which a quarter are white and the remainder are black.

There are, before the drawings, 11 possible hypotheses concerning the contents of the urn. They range from 0 whites and 10 blacks to 10 whites and 0 blacks, as listed in the two left-hand columns of Table I and shown in Fig. 1. The probability in favor of each of

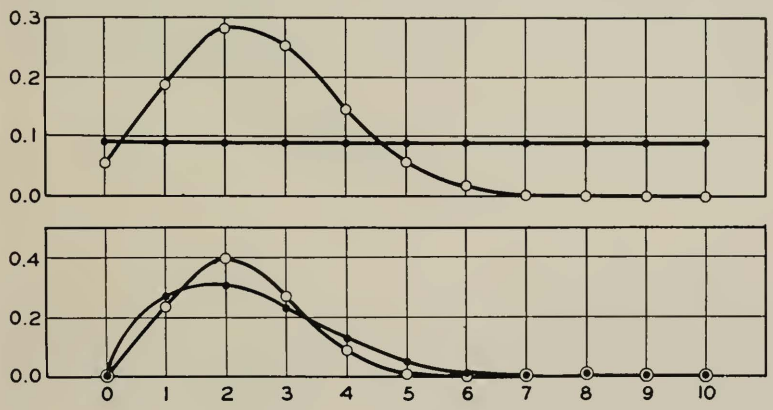


Fig. 1. The upper curve shows two different assumptions concerning the *a priori* probabilities, while the lower pair shows the *a posteriori* probabilities. In both cases the dots refer to the hypothesis of uniform *a priori* probability while the circles refer to the assumption that the urn itself is a random sample from a large stock of which one fourth of the balls are white.

these hypotheses is the "*a priori* existence probability" in favor of the hypotheses, and is represented by the symbol $w(X)$, X referring to the number of white balls assumed to be in the urn.

Under assumption "a" (Case 1) each hypothesis has a probability of 1/11 or .090909. Under assumption "b" (Case 2) the probability that the urn contains X whites and $10 - X$ blacks is the binomial term $\binom{10}{X}\left(\frac{1}{4}\right)^{X}\left(\frac{3}{4}\right)^{10-X}$.

TABLE I

Contents		Existence Prob. $w(X)$		Prod. Prob.	*A Posteriori* Prob. P_X	
"X" Wh.	Bl.	Case 1	Case 2	p_X	Case 1	Case 2
0	10	.090909	.056314	.000000	.000000	.000000
1	9	.090909	.187712	.500000	.272727	.237305
2	8	.090909	.281568	.555556	.303030	.395509
3	7	.090909	.250282	.416667	.227272	.263671
4	6	.090909	.145998	.238095	.129870	.087889
5	5	.090909	.058399	.099206	.054112	.014650
6	4	.090909	.016222	.023810	.012987	.000876
7	3	.090909	.003090	.000000	.000000	.000000
8	2	.090909	.000386	.000000	.000000	.000000
9	1	.090909	.000029	.000000	.000000	.000000
10	0	.090909	.000001	.000000	.000000	.000000

In the column headed p_X we give the productive probabilities for both cases. These are the probabilities that five drawings from an urn whose contents were as given by the corresponding hypothesis would yield the observed one white ball and four black. These are zero in the case of $X = 0$, 7, 8, 9 and 10 since urns so constituted could not have given the observed drawings.

For the other cases, the productive probability is the ratio

$$p_X = \frac{\binom{X}{1}\binom{10 - X}{4}}{\binom{10}{5}}.$$

In this expression the denominator is the total number of combinations of 10 balls taken five at a time, and the numerator is the number of ways of selecting one out of X white balls and four out of the remaining $10 - X$ black balls. These figures are tabulated in Table I under the heading p_X.

We now have all of the component parts of our problem under the two different assumptions "a" and "b." It only remains to apply "Bayes' Rule."[2] Now the generalized Bayes Rule tells us that the *a posteriori* probability, P_X, in favor of an hypothesis *after* the drawings

[2] This rule was first enunciated by an English cleric, Bayes by name, in a memoir in *Philosophical Transactions* for 1763. It was generalized by Laplace in 1812 to cover cases not equally likely.

have been made and taking account of the *a priori* information is given by the ratio

$$P_X = \frac{w(X)p_X}{\sum w(X)p_X},$$

the summation in the denominator being extended over all possible cases.

The numerical values of this ratio are shown in the last two columns of Table I, corresponding to the two assumptions in our problem and also by the circles in Fig. 1. That we should have a different set of results corresponding to the different assumptions is to be expected. It is interesting, however, that the difference in this case is by no means great as Fig. 1 brings out.

If after each drawing we had replaced the ball drawn, we would have used for the productive probability p_X the binomial term

$$p_X = \binom{5}{1}\left(\frac{X}{10}\right)^1\left(\frac{10 - X}{10}\right)^4$$

since the successive drawings would not have changed the relative constitution of the urn. The same would also be true if the urn contained an indefinitely large number of balls with the same relative proportions of black and white.

Now if we agree that a white ball corresponds to a defective item and a black ball to an acceptable item, we are immediately able, by the use of these fundamental principles of *a posteriori* probability, to write the general basic formal relation

$$W(X_1, X_2) = \frac{\sum_{X=X}^{X=X_2} w(X)\binom{X}{c}\binom{N-X}{n-c}}{\sum_{X=c}^{X=N-n+c} w(X)\binom{X}{c}\binom{N-X}{n-c}}. \qquad (1)^3$$

As we have just indicated, the troublesome element in this formula is the function $w(X)$ to which, in many practical problems, it is difficult to assign any particular numerical values. In order to proceed further, therefore, without detailed consideration of various specific engineering problems we are forced to make some rather general assumptions concerning the nature of the function $w(X)$.

Case I

One of the most natural assumptions to make when no knowledge exists to the contrary is that $w(X)$ is a constant within that range

[3] It should be noted that in his original treatment of this formula Molina used S instead of Σ as the symbol for summation on account of the fact that finite integration entered into his analysis. Since in this presentation we are dealing only with summation, we shall use the commoner form Σ to denote summation.

of values of X which essentially affects the value of the denominator of (1). This assumption may seem at first glance rather arbitrary and wide of the mark, especially since the range of values which essentially affects the value of the denominator in (1) depends on the value of c obtained from the sample. However, if the sample is reasonably large, consisting of 100 items or more, and the proportion of defectives observed is small, say 10 per cent or under, the probability that universes having a proportion of defectives widely different from the one observed would yield such results is so small that a wide range of assumptions concerning the *a priori* probability of such universes existing makes very little change in the final result.

Applying, then, this assumption analytically to the basic formula (1) we obtain the simpler formulæ

$$W(X_1, X_2) = \frac{\sum_{X=X_1}^{X=X_2} \binom{X}{c}\binom{N-X}{n-c}}{\sum_{X=c}^{X=N-n+c} \binom{X}{c}\binom{N-X}{n-c}} = \frac{\sum_{X=X_1}^{X=X_2} \binom{X}{c}\binom{N-X}{n-c}}{\binom{N+1}{n+1}} \quad (2)$$

and by means of a transformation outlined in the Appendix we obtain, from (2),

$$W(X_1, X_2) = \frac{\sum_{t=0}^{t=c}\left[\binom{X_1}{t}\binom{N+1-X_1}{n+1-t} - \binom{X_2+1}{t}\binom{N-X_2}{n+1-t}\right]}{\binom{N+1}{n+1}} . \quad (2a)$$

Formula (2*a*) is the one embodied in the paper referred to in footnote 1. While apparently less simple than (2), it is actually easier to compute when c is less than the range $X_2 - X_1$.

When in (2*a*) we set $X_1 = c$ and $X_2 = X$ the resulting formula

$$W(c, X, n, N) = 1 - \frac{\sum_{t=0}^{t=c} \binom{X+1}{t}\binom{N-X}{n+1-t}}{\binom{N+1}{n+1}}, \quad (3)$$

which is at the basis of our computational work, shows explicitly certain properties which are not apparent in (2). Various analytical transformations and approximations based on this formula lead to several interesting extensions which are discussed in the Appendix. We shall leave these phases of the problem for the present, however, and discuss the results of the calculations which have been made as presented on the attached charts.

Charts A

Charts A have been prepared by means of exact formula (3) to show, for universes $N = 300$, 500, 700 and 900 from which samples, n, of various indicated sizes are assumed to have been drawn, the probability or "weight" $W(c, X)$ as ordinate versus X as absicissa for various values of c as indicated by the solid curves so designated. The dotted curves crossing these solid curves show the weight indicated by various values of the difference "d" between the percentage observed defective and the percentage assumed defectives in the universe.

As examples illustrating the interpretation of Charts A consider the following:

Example 1: From a universe of $N = 700$ items a random sample $n = 300$ items has shown $c = 3$ or 1 per cent defectives. What is the probability or weight to be associated with the hypothesis that the universe contains not more than $X = 14$ or two per cent defectives? From the A Chart corresponding to $N = 700$ and $n = 300$ we find the $c = 3$ curve (shown heavy because it is an even per cent of the sample $n = 300$). On this curve corresponding to an abscissa of $X = 14$ we read our desired result as the ordinate $W = .94$. We note that this is also a point on the $d = 1$ per cent dotted curve since $100(X/N - c/n)$ per cent = 1 per cent.

Example 2: We are going to make a sample of $n = 199$ items out of a universe of $N = 500$ items and wish the weight or probability to be .9 or better that the universe does not contain more than five per cent defective items. What is the maximum number of defective items that we may tolerate in our sample? Now five per cent of $N = 500$ is $X = 25$. Corresponding to an abscissa $X = 25$ and an ordinate $W = .9$ we locate a point which lies between the $c = 6$ and $c = 7$ curves. We could, therefore, accept the lot provided the sample showed six or less defectives, or three per cent or less defectives.

These Charts A are fundamental in nature, and involve the five variables, N, n, X, c and W. The formula by means of which they were computed is exact on the basis of the assumptions. Such errors or irregularities as may appear to exist in them are of negligible practical importance in view of the nature of the assumptions made, and are mainly due to the difficulties in drafting such a family of curves.

Naturally a function involving several variables may be represented graphically in many different ways, some of which may be more convenient than others to use in connection with various practical problems. One of the restrictions often encountered in practical

problems is that the weight W shall not be less than some specified figure which may be considered to give us the desired degree of confidence in the efficacy of our sampling procedure in weeding out defective lots. Charts B and C are drawn up on the basis of three such specified figures which are of practical interest, $W = .75$, $W = .9$, and $W = .99$. Such restrictions enable us to show, without the large amount of labor which would be required without them, the results of calculations for a wider range of the other variables.

Chart B

Chart B shows roughly for the proportion of observed defectives $c/n = .01$, .04, and .07, the proportion of defectives in the universe which we may expect not to exceed with weights $W = .75$, .9 and .99 for various values of the sample n as abscissa and for $N = 300$, 500, 700, 900 and also the limit approached as N becomes infinite. This form of presentation serves to relate the present material to the earlier charts which accompanied the earlier article already mentioned as having appeared in the *Bell System Technical Journal* for January, 1924, and shows how with a given size of sample n and a given proportion of defectives observed, the larger the value of the universe N, the larger the variation which may be expected with any given degree of probability. As would be expected, we also see that when the size of the sample approaches the size of the universe, the range of uncertainty approaches 0 and our sample inspection becomes a complete inspection.

It will be noted that, up to the present point, we have not considered cases for $N > 1{,}000$. The exact formulæ become rather troublesome to compute for these larger values of N. Fortunately, however, various approximate methods outlined in the Appendix become sufficiently accurate to be of service in these cases.

Charts C

We have, therefore, by their aid when $N > 1{,}000$, prepared the Charts C which we believe will cover a rather wide range of the variables with sufficient precision to be of considerable practical value. The points shown by dots are believed to be accurate to the degree to which they are readable on the chart. For intermediate values and for other values of the trouble limit the discrepancies are indicated on the charts. One of these charts corresponds to each of the three following weights, $W = .75$, $W = .9$, and $W = .99$. As abscissa we show the per cent sample, $100\ n/N$. The ordinate scale is proportional to the number of items n in the sample. The same

proportionality factor K enters also in the ratio X/N which we designate as the trouble limit. We shall later discuss the purpose of this factor K in more detail. The understanding of the charts will be simplified, however, if we consider the case for $K = 1$ in which the charts become direct reading for the case of a trouble limit $X/N = .01$.

The values of c, the number of defective items observed in the sample, are shown as a family of curves marked $c = 0$, $c = 1$, $c = 2$, etc., sloping downward from left to right. Any point on the $c = 5$ curve, for example, on the Chart C for weight $W = .9$ shows the corresponding values of n as ordinate and n/N as abscissa which are necessary in order that this number of defectives may be accepted with a degree of assurance [4] indicated by $W = .9$ that the true proportion of defectives in the universe N is not greater than .01.

It will be readily noted that for every value of the universe N, there may be drawn a diagonal straight line through the origin whose ordinate for an abscissa of 100 per cent sample is equal to $n = N$. Certain representative N lines are drawn in on the charts in this manner, and as many more could be inserted as desirable. Thus, for a constant value of W and a constant value of X/N we have provided on Charts C a ready means of determining the relationships which must exist between the remaining variables N, n, and c.

As an example of the use of these charts for the case where $K = 1$, i.e., for $X/N = .01$, consider the following:

Example 3: In a sample of $n = 900$ out of a universe $N = 3{,}000$, what is the maximum number of defectives c that we may accept with an assurance of $W = .9$ or better that the true proportion of defectives in the universe is not greater than .01?

Referring to the Charts C for $W = .9$ and considering $K = 1$, we locate the point corresponding to an abscissa of 100 n/N per cent $= 90{,}000/3{,}000 = 30$ per cent, and an ordinate $n = 900$. We find that this lies on the diagonal straight line marked $N = 3{,}000$ K as it should and that it also lies between the $c = 5$ and $c = 6$ curves. From this we may infer that we may accept five defectives but not six in the above case.

We shall now proceed to explain the significance of the factor K and the cross-hatched areas beneath the $c = 0$, 5, 10, 15, etc., curves. The purpose of these features is to extend the application of Charts C to values of X/N other than .01. It may be noted from the mathematical analysis or from actual plotting of charts similar to Charts C, but for different values of X/N, that the general shape and spacing of the curves remains practically unchanged for any given value of W.

[4] This statement is not strictly true when we are dealing with non-integral values of X. In such cases the weights W shown on the Charts C are slightly too high.

In other words, the value of $W(c, X)$ depends mainly on the ratio n/N, and the values of X and c, and only in a secondary way on the absolute values of n and N. This being the case, if we make a given per cent sample of two different universes N and KN, the number of defectives c which we may allow in our sample out of the first universe N in order that our weight W may have a given value, .9 say, for the true proportion of defectives in this universe to be not greater than .01 is practically the same as the value of c that we may allow in the sample out of the second universe KN for the same weight W and a proportion of defectives $.01/K$. For values of $K > 1$ there is no appreciable change introduced in the location of the c curves on Charts C. For values of $K < 1$, some error is made. The magnitude of this error is indicated by the cross-hatched bands on the $c = 0$, 5, 10, 15, etc., curves. The lower boundaries of these bands were calculated to show the magnitude of the error introduced for the corresponding values of c when $K = .1$. The upper boundaries of these areas correspond to values of $K \geqq 1$. For other values of c only the upper boundaries of the corresponding bands are shown, the lower boundaries being easily deducible by visual interpolation to a sufficient degree of approximation for most practical purposes.

As examples which may serve to illustrate this sort of application of Charts C consider the following:

Example 4: A sample of $n = 5{,}000$ items has been drawn out of a universe of $N = 20{,}000$ items and $c = 15$ defectives were observed. May we assume with a weight $W = .9$ or more that the true proportion of defectives or trouble limit X/N is .005?

Here $.01/K$ is to equal .005 for our charts to apply. Therefore, $K = 2$. Our sample $n = 500 = 2{,}500K$ and our per cent sample is $100\ n/N = 25$ per cent. Corresponding then to an abscissa of 25 per cent and an ordinate of $2{,}500K$ on the $W = .9$ chart we locate a point between the $c = 19$ and $c = 20$ curves. We could have allowed, therefore, $c = 19$ defectives at the desired weight and trouble limit. Since we observed a smaller number of defectives than was allowed, our weight W is therefore greater than .9. As a matter of fact it is practically only slightly less than .99 as appears from the $W = .99$ chart when utilized in a corresponding manner.

Example 5: As our next example we shall attempt to determine what is the trouble limit which corresponds with $W = .9$ to the results of the sample of Example 4. On the $W = .9$ chart corresponding to an abscissa of 25 per cent we read from the $c = 15$ curve an ordinate of $2{,}015K$. But this must be our sample $n = 5{,}000$. We, therefore, determine K from the equation $2{,}015K = 5{,}000$ which gives $K = 2.48$. Hence, our corresponding trouble limit is

$$\frac{.01}{K} = \frac{.01}{2.48} = .0040.$$

So far our values of K have been greater than unity, so we have not had to consider our cross-hatched bands at all. In our next example we shall remedy this defect.

Example 6: What number of defective items c may be allowed in a sample of 200 items out of a universe of 500 items so that $W \geq .9$ corresponding to a trouble limit $X/N = .08$. Here $.01/K = .08$ $\therefore K = .125$. Our ordinate, therefore, is $200 = 1{,}600K$ and our abscissa is $200/500 \times 100 = 40$ per cent sample. The point corresponding to this on the $W = .9$ chart lies just below the $c = 12$ curve indicating at first glance that we could not accept 12 defectives in such a case. However, we note that $K = .125$ should be near the lower boundary of our cross-hatched band for $c = 12$ if such a band had been drawn in. From an inspection of the widths of the bands for $c = 10$ and $c = 15$ we correctly infer that our point determined by the 40 per cent sample and $1{,}600K$ would lie well within this band, and that after all we could accept 12 defectives in the example in question.

This example has been included merely to illustrate the interpretation of the bands shown on Charts *C*. It may be anticipated that in many if not most of the practical engineering problems only the upper boundaries of the bands need be used to obtain a degree of accuracy commensurate with the precision of the results desired and the applicability of the basic assumptions concerning randomness and the form of the *a priori* existence probability $w(X)$.

If it should be desired to extend the range of these charts to cover values of W other than those shown, this may be done by means of the methods outlined in the mathematical analysis, the particular method to be used depending on the degree of precision required.

The preceding pages have contained an outline of some of the theory and results based on the assumption that, within a range at least, all possible values of X, the unknown number of defectives, were *a priori*, that is, before the sample in question was made, equally likely. This assumption we mentioned as appropriate to consider in case we have no information to the contrary. The results may be also applicable to certain cases where we do have some information of a general sort, but which it is difficult to express analytically. However, it is by no means the only reasonable assumption to make concerning the form of $w(X)$ as it enters into the basic formula (1).

Case II

Another assumption is suggested by the following considerations which enter into many of the standard works on probability theory. Assume that the lot or universe in question was itself drawn at random from an extremely large stock or major universe in which the proportion of defective items was p. Under these conditions the *a priori* probability $w(X)$ that our universe of N items would contain exactly X defective items would be given by the expression

$$w(X) = \binom{N}{X} p^X (1 - p)^{N-X}.$$

Using this expression for $w(X)$ in the fundamental formula (1), we obtain, by a process given in detail under the heading Appendix, the formula

$$W(X_1, X_2) = \sum_{X=X_1}^{X=X_2} \binom{N-n}{X-c} p^{X-c} (1 - p)^{N-n-X+c},$$

which for $X_1 = c$ and $X_2 = X$ reduces to

$$W(c, X) = \sum_{t=0}^{X-c} \binom{N-n}{t} p^t (1 - p)^{N-n-t},$$

which is precisely the expression for the *a priori* probability that the remaining $N - n$ items which we did not inspect contain not more than the $X - c$ defectives which together with the c we have observed would assure us of a satisfactory universe.

In this form $W(c, X)$ turns out to be a simple binomial which, when $N - n$ is large and p is small, may be reasonably approximated by the Poisson Exponential Binomial Limit for which extensive curves and tables already exist[5] and will, therefore, not be included in this article.

In order to make practical use of the results of this assumption, we must have some knowledge of the appropriate value of the factor p in any given case. This factor should measure the probability that an item, selected at random, will, on inspection, prove to be defective. If a large number of tests have been made in the past on similar items, prepared by essentially the same process, the ratio of the total defectives observed to total items inspected in such tests may be a reasonable figure to use for p. In the case of many manufactured articles such a ratio ought not to be very difficult to obtain. In certain cases it might be necessary to allow for such factors as

[5] See article, "Probability Curves Showing Poisson's Exponential Summation," by George A. Campbell, *Bell System Technical Journal*, January, 1923.

trend, improved process of manufacture, changes in personnel and so on. In the final sampling of such complicated equipment as is involved in a completely installed telephone central office it may be necessary to take account also of breakage and such troubles due to shipping and setting up of the equipment which may introduce marked deviations from average conditions.

Such general considerations as these should determine whether or not we can safely assume any given value for p, and if so, what value. It will be evident on a little consideration that the assumed value for p need not be extremely precise for many practical applications.

Concerning the restrictions on the function giving the *a priori* probabilities, $w(X)$, it may be well to point out that this function is only defined for the positive integral values of X such that $0 \leq X \leq N$. Moreover, since probabilities are essentially positive, it cannot be negative for any of these values of X. Also since the composition of the lot is certainly comprised in this range of values of X, one has

$$\sum_{0}^{N} w(X) = 1.$$

The questions raised concerning the form of $w(X)$ are of particular importance in connection with the economic phases of sampling, that is, the relative costs of having satisfactory lots rejected and unsatisfactory lots accepted by the sampling process. These costs are, of course, dependent on the frequency with which given proportions of defects occur in the lots in practice, and a detailed consideration of these would itself warrant a separate treatment.

It is felt that the general methods outlined in this treatment, while not sufficiently detailed for immediate practical application to many of the problems in sampling of attributes, will nevertheless serve as a satisfactory basis for further work of a more specific nature.

APPENDIX

Case I: Assuming $w(X)$ is a constant and noting that [6]

$$\sum_{X=c}^{X=N-n+c} \binom{X}{c}\binom{N-X}{n-c} = \binom{N+1}{n+1},$$

the fundamental formula

$$W(X_1, X_2) = \frac{\sum_{X=X_1}^{X=X_2} w(X)\binom{X}{c}\binom{N-X}{n-c}}{\sum_{X=c}^{X=N-n+c} w(X)\binom{X}{c}\binom{N-X}{n-c}} \tag{1}$$

[6] Netto, "Lehrbuch der Kombinatorik," p. 15, Eq. 11.

gives us, as one computing formula,

$$W(X_1, X_2) = \frac{\sum_{X=X_1}^{X=X_2} \binom{X}{c}\binom{N-X}{n-c}}{\binom{N+1}{n+1}}, \tag{2}$$

which is fairly manageable so long as the range X_1 to X_2 is not too great and we have tables for the logarithms of the factorials involved. When $X_2 - X_1$ is large compared with c, one may use the equivalent formula

$$W(X_1, X_2) = \frac{\sum_{t=0}^{t=c} \binom{X_1}{t}\binom{N+1-X_1}{n+1-t} - \binom{X_2+1}{t}\binom{N-X_2}{n+1-t}}{\binom{N+1}{n+1}}. \tag{2a}$$

This transformation may, as Molina has shown, be effected as follows: [7]

$$\sum_{X=X_1}^{X=X_2}\binom{X}{c}\binom{N-X}{n-c} = \sum_{X=X_1}^{X=N-n+c}\binom{X}{c}\binom{N-X}{n-c} - \sum_{X=X_2+1}^{X=N-n+c}\binom{X}{c}\binom{N-X}{n-c}.$$

Now

$$\begin{aligned}\sum_{X=X_2+1}^{X=N-n+c}\binom{X}{c}\binom{N-X}{n-c} &= \sum_{X=X_2+1}^{X=N-n+c}\binom{N-X}{n-c}\left(\sum_{t=0}^{t=c}\binom{X_2+1}{t}\binom{X-X_2-1}{c-t}\right)\\ &= \sum_{t=0}^{t=c}\binom{X_2+1}{t}\left(\sum_{X=X_2+1}^{N-n+c}\binom{N-X}{n-c}\binom{X-X_2-1}{c-t}\right)\\ &= \sum_{t=0}^{t=c}\binom{X_2+1}{t}\binom{N-X_2}{n+1-t}.\end{aligned}$$

Likewise

$$\sum_{X=X_1}^{X=N-n+c}\binom{X}{c}\binom{N-X}{n-c} = \sum_{t=0}^{t=c}\binom{X_1}{t}\binom{N+1-X_1}{n+1-t}.$$

If in (2a) we let $X_1 = c$ and $X_2 = X$, we obtain

$$W(c, X) = 1 - \frac{\sum_{t=0}^{t=c}\binom{X+1}{t}\binom{\overline{N+1} - \overline{X+1}}{n+1-t}}{\binom{N+1}{n+1}}, \tag{3}$$

from which we may compute $W(X_1, X_2)$ from the equation

$$W(X_1, X_2) = W(c, X_2) - W(c, X_1 - 1).$$

[7] See also Netto, "Lehrbuch der Kombinatorik," p. 12, Eq. 6; p. 15, Eq. 11.

A direct interpretation of the expression for $W(c, X)$ gives us the following interesting

Theorem A: The *a posteriori* probability that a universe of N items contains not more than X defectives when c defectives have resulted from a random sample of n items is equal to the *a priori* probability of obtaining at least $c + 1$ defectives in a random sample of $n + 1$ items from a universe of $N + 1$ items of which exactly $X + 1$ are defective. This theorem assumes that *a priori* all values of X are equally likely.

Writing $s = X + 1 - t$, we obtain from (3)

$$1 - W(c, X) = 1 - \frac{\sum_{s=0}^{s=X-c} \binom{X+1}{s} \binom{\overline{N+1} - \overline{X+1}}{N-n-s}}{\binom{N+1}{N-n}}. \quad (4)$$

The right-hand side of this equation is exactly what would have been obtained directly from (3) if we had been dealing with a sample of $N - n - 1$ instead of n and had observed $X - c$ defectives instead of c, since the particular symbol chosen for the variable of summation is immaterial.

This fact, which follows immediately from physical consideration of the equivalent *a priori* problem of Theorem A, may be stated as

Theorem B: If we calculate the probability W that a universe of N items contains not more than X defectives when a sample of n has shown exactly c defectives, then $1 - W$ is the probability that a universe of N items contains not more than X defectives when a sample of $N - n - 1$ has shown exactly $X - c$ defectives.

In making extensive calculations, this relation will serve to cut down the amount of computation considerably, as each calculated value of W may be made to do double duty. For a single calculation either (3) or (4) may be used depending on which involves the shorter summation.

Another interesting relation also appears when we note that

$$\frac{\binom{X+1}{t}\binom{N-X}{n+1-t}}{\binom{N+1}{n+1}} = \frac{\binom{n+1}{t}\binom{N-n}{X+1-t}}{\binom{N+1}{X+1}},$$

which may be proved simply by cross multiplication of the combination factors, writing them in terms of factorials. From this we see that

$$W(c, X) = 1 - \frac{\sum_{t=0}^{t=c} \binom{n+1}{t} \binom{\overline{N+1} - \overline{n+1}}{X+1-t}}{\binom{N+1}{X+1}}. \quad (5)$$

If we compare the right-hand sides of (3) and (5), we see that n and X have simply been interchanged, which proves the rather interesting

Theorem C: The probability $W(c, X)$ that a universe of N items does not contain more than X defectives when a sample of n has shown exactly c defectives is equal to the probability $W(c, n)$ that a universe of N items does not contain more than n defectives when a sample of X has shown exactly c defectives.

Thus equations (3), (4) and (5) taken together show that we may make three different interpretations of a single calculation.

Up to this point, all of the analysis has been *exact* on the basis of the fundamental assumptions. We may now proceed with advantage to consider some approximate relationships which have been for some years of service in the calculation of practical curves and tables for cases where the values of N, n, or X were too great to be handled conveniently by means of the exact formulæ.

Now consider in formula (3) a single term, π_t say, where

$$\pi_t = \binom{X+1}{t} \frac{\binom{N-X}{n+1-t}}{\binom{N+1}{n+1}}$$

$$= \binom{X+1}{t} \frac{\left(\frac{(n+1)!}{(n+1-t)!}\right)\left(\frac{(N-n)!}{(N-n-X-1+t)!}\right)}{\left(\frac{(N+1)!}{(N-X)!}\right)}$$

$$= \binom{X+1}{t}\left(\frac{n}{N}\right)^t\left(1-\frac{n}{N}\right)^{X+1-t} \cdot F(N, n, X, t), \qquad (6)$$

where the form of the function F is to be determined.

To facilitate the consideration of this function we may split it up into three similar parts as follows:

$$F(N, n, X, t) = \frac{\varphi(n+1, t-1) \cdot \chi(N-n, X-t)}{\varphi(N+1, X)},$$

where

$$\varphi(n+1, t-1) = \left(1+\frac{1}{n}\right)\left(1\right)\left(1-\frac{1}{n}\right) \cdots \left(1-\frac{t-2}{n}\right),$$

$$\varphi(N+1, X) = \left(1+\frac{1}{N}\right)\left(1\right)\left(1-\frac{1}{N}\right) \cdots \left(1-\frac{X-1}{N}\right),$$

$$\chi(N-n, X-t) = \left(1\right)\left(1-\frac{1}{N-n}\right) \cdots\cdots\cdots \left(1-\frac{X-t}{N-n}\right).$$

Recalling that

$$\log (1 + X) = X - \frac{1}{2}X^2 + \frac{1}{3}X^3 - \frac{1}{4}X^4 + \cdots,$$

which converges for $X^2 < 1$,

we have

$$\log \varphi(n + 1, t - 1) = \log\left(1 + \frac{1}{n}\right) - \sum_{y=0}^{t-2}\sum_{r=1}^{\infty}\frac{1}{r}\left(\frac{y}{n}\right)^r,$$

$$\log \varphi(N + 1, X) = \log\left(1 + \frac{1}{N}\right) - \sum_{y=0}^{X-1}\sum_{r=1}^{\infty}\frac{1}{r}\left(\frac{y}{N}\right)^r,$$

$$\log \chi(N - n, X - t) = -\sum_{y=0}^{X-t}\sum_{r=1}^{\infty}\frac{1}{r}\left(\frac{y}{N - n}\right)^r,$$

whence

$$\log F(N, n, X, t) = \log\left(1 + \frac{1}{n}\right) - \log\left(1 + \frac{1}{N}\right) + \sum_{y=0}^{X-1}\sum_{r=1}^{\infty}\frac{1}{r}\left(\frac{y}{N}\right)^r - \sum_{y=0}^{t-2}\sum_{r=1}^{\infty}\frac{1}{r}\left(\frac{y}{n}\right)^r - \sum_{y=0}^{X-t}\sum_{r=1}^{\infty}\frac{1}{r}\left(\frac{y}{N - n}\right)^r$$

Neglecting terms of the second order in $1/n$, $1/N$ and $1/(N - n)$, we have as an approximation

$$\log F(N, n, X, t) \doteq -\frac{1}{2}\left(\frac{t^2 - 3t}{n} + \frac{X^2 - 2Xt + t^2 + X - t}{N - n} - \frac{X^2 - X - 2}{N}\right).$$

If now we select as our value of t, $t = (n/N)X$, we have $\log F(N, n, X, t) \doteq (X - 2)/2N$ which is > 0; if $X > 2$,

$$F(N, n, X, t) \doteq e^{(X-2)/2N},$$

which gives us as an approximate value for the maximum term π_t, where $t = (n/N)X$,

$$\pi_t' = \left(\frac{X + 1}{t}\right)\left(\frac{n}{N}\right)^t\left(1 - \frac{n}{N}\right)^{X+1-t} e^{(X-2)/2N}. \tag{7}$$

Having this term, it is a simple matter to calculate the other terms necessary for evaluating $W(c, X)$ by means of the exact equations

$$\pi_{t+1} = \pi_t\left(\frac{n - t + 1}{t + 1} \cdot \frac{X - t + 1}{N - n - X + t}\right), \tag{8}$$

$$\pi_{t-1} = \pi_t\left(\frac{t}{n - t + 2} \cdot \frac{N - n - X + t - 1}{X - t + 2}\right). \tag{9}$$

Due to the reciprocal relationship between n and X, we may obtain in a similar manner

$$\binom{n+1}{t}\frac{\binom{N-n}{x+1-t}}{\binom{N+1}{X+1}}=\binom{n+1}{t}\left(\frac{X}{N}\right)^{t}\left(1-\frac{X}{N}\right)^{n+1-t}e^{(n-2)/2N}, \quad (10)$$

when $t = (X/N)n$.

It is by means of these relationships that we have calculated the cases for $N > 1{,}000$ as shown on Charts C and feel that the precision obtained is rather better than would have resulted from using formula (6) for all values of t and assuming $F(N, n, X, t) = 1$. However, for suitable ranges of the variables involved, the formula resulting from this procedure

$$W(c, X) \doteq 1 - \sum_{t=0}^{t=c}\binom{X+1}{t}\left(\frac{n}{N}\right)^{t}\left(1-\frac{n}{N}\right)^{X+1-t} \quad (11)$$

would be a fairly good approximation. This is simply part of the well-known binomial expansion and is far simpler to compute than the more precise formulæ, although by no means easy at that.

We may draw several interesting practical conclusions, however, from formula (11). For instance, we may note that as n/N approaches 0 and X becomes infinite in such a way that the product $(X + 1)(n/N)$ remains constant and equal to the average a, we have the familiar Poisson Exponential Binomial Limit

$$W(c, X) = 1 - \sum_{t=0}^{c}\frac{a^{t}e^{-a}}{t!},$$

where $a = (X + 1)(n/N)$.

In addition we note from formula (11) that, for small values of X/N, the variable N enters into the formula only in the ratio n/N. From this we deduce the fact, borne out by independent calculations, that by means of the proper use of a proportionality factor K applied directly to n and N and inversely to X/N we may extend the Charts C to care for values of $X/N \leqq .1$ to a very good degree of accuracy and with considerable saving in space and computational labor.

By the reciprocal relationship between X and n as shown in exact formulæ (3) and (5), we obtain

$$W(c, X) \doteq 1 - \sum_{t=0}^{t=c}\binom{n+1}{t}\left(\frac{X}{N}\right)^{t}\left(1-\frac{X}{N}\right)^{n+1-t}, \quad (12)$$

which differs only in form from equation (3) of Molina's paper [8] on

[8] Footnote 1.

the infinite universe case. Formula (12) does not give the same results as (11) as it is most exact when n/N is small and becomes absolutely exact in the limiting case of an infinite universe where $n/N \doteq 0$. This formula also approaches the Poisson Limit, in this case as X/N approaches 0 and $n + 1$ becomes infinite in such a way that the product $(n + 1)(X/N)$ remains constant and equal to α, say.

The Poisson Limit, for the case of an infinite universe, was given by Molina in the Appendix to the article in the *Bell System Technical Journal* of January, 1924, already mentioned in this memorandum.

Another point of interest is brought out when we note that in the limiting form of (12) the Poisson gives us

$$W(c, X) \doteq \sum_{t=c+1}^{\infty} \frac{a^t e^{-a}}{t!}, \quad a = \frac{X \cdot n}{N},$$

and for another pair of values of W and X

$$W_1(c, X_1) = \sum_{t=c+1}^{\infty} \frac{a_1^t e^{-a_1}}{t!}, \quad a_1 = \frac{X_1 \cdot n}{N}.$$

Thus from properly chosen Poisson curves or tables we may obtain the ratio $X_1/X \doteq a_1/a$ which corresponds to the observed value of c and the desired values of W and W_1. This ratio in exact formulæ is a function of N, n, and X also, but for many problems involving small values of n/N and X/N the degree of approximation furnished by this limiting form is fairly satisfactory and still further reduces the amount of labor necessary in extending approximate results to practice.

The sort of procedure we have just been discussing may be facilitated by means of a chart on which we show as abscissæ values of c and as ordinates values of the ratio of X_1/X which corresponds to various values of W as shown by various curves and a specified value of W_1, say $W_1 = .9$. Such a chart would enable us to interpret roughly a given Chart C for $W = .9$ in terms of other values of W. For precise work this procedure is not to be recommended, and, therefore, no charts of the nature just described are included herein.

Approximations to the binomial other than the Poisson have been discussed in many of the texts. In particular, for values of p in the neighborhood of $\frac{1}{2}$, the well-known Laplace-Bernoulli integral

$$\frac{1}{\sqrt{\pi}} \int_a^b e^{-t^2} dt$$

will serve as an approximate value for W_1 where the limits a and b

are functions of N, n, X, and c. This approximation is not so suitable, however, for most telephone sampling problems in which the proportion of defectives may be assumed in general to be far smaller than $\frac{1}{2}$.

We shall now proceed to discuss a few points concerning the analysis of Case II in which instead of assuming $w(X)$ constant we assumed it to be of the form

$$w(X) = \binom{N}{X} p^X(1-p)^{N-X}.$$

Combining this expression for $w(X)$ with the term $\binom{X}{c}\binom{N-X}{n-c}$ which appears in the basic formula (1), we have

$$\frac{X!}{c!(X-c)!} \cdot \frac{(N-X)!}{(n-c)!(N-X-n+c)!} \frac{N!}{X!(N-X)!} p^X(1-p)^{N-X}$$
$$= \binom{N}{n} \cdot \binom{n}{c} p^c(1-p)^{n-c} \cdot \left(\binom{N-n}{X-c} p^{X-c}(1-p)^{N-n-X+c} \right).$$

Since only the factors in brackets involve the variable of summation X, the remainder of this expression will cancel out in numerator and denominator, leaving us with

$$W(X_1, X_2) = \frac{\sum_{X=X_1}^{X=X_2} \binom{N-n}{X-c} p^{X-c}(1-p)^{N-n-X+c}}{\sum_{X=c}^{X=N-n+c} \binom{N-n}{X-c} p^{X-c}(1-p)^{N-n-X+c}}$$

as the resulting form for (1) with this assumption for $w(X)$.

It may be noted that the summation in the denominator above is a complete binomial $(p+q)^{N-n}$ and as such equals unity, so

$$W(X_1, X_2) = \sum_{X=X_1}^{X=X_2} \binom{N-n}{X-c} p^{X-c}(1-p)^{N-n-X+c},$$

where p is assumed to be the *a priori* probability of a defective item as determined from reliable information concerning conditions under which the items are prepared.

As before when $X_1 = c$ and $X_2 = X$ we have

$$W(c, X) = \sum_{t=0}^{t=X-c} \binom{N-n}{t} p^t(1-p)^{N-n-t}.$$

We may be willing in certain cases to admit the binomial form for

$w(X)$ without being able or willing to assign any single value to p. In such cases we may, however, proceed to make assumptions concerning the probability that p has a given value. Let

$$s(X, p) = f(p)\binom{N}{X} p^X(1 - p)^{N-X};$$

then

$$w(X) = \int_0^1 s(x, p)dp = \binom{N}{X}\int_0^1 f(p)p^X(1 - p)^{N-X}dp,$$

where

$$\int_0^1 f(p)dp = 1 \quad \text{and} \quad \sum_{X=0}^{N} w(X) = 1.$$

Suppose we assume $f(p)$ constant for all values between 0 and 1; we have

$$w(X) = \binom{N}{X}\int_0^1 p^X(1 - p)^{N-X}dp = \frac{1}{N + 1},$$

which we note to be a constant which assigns to all of the $N + 1$ possible *a priori* hypotheses concerning X an equal weight. This pair of assumptions in Case II amounts, therefore, to the same thing analytically as the assumption of Case I.

Any number of possible hypotheses concerning $f(p)$ might be made. Some of these would complicate the analysis considerably, others might be carried through fairly simply. One of these hypotheses might fit one class of physical problems, another some other class. To consider these all in detail in this paper would be outside of the scope of a general treatment. The methods outlined here would, however, hold for such extensions. Such difficulties as might be encountered would be of an analytical rather than a logical nature.

In closing, the author wishes to express his appreciation to his numerous friends and associates in the Bell System, whose suggestions and cooperation have been of material assistance in the preparation of this work, and particularly the work of Miss Nelliemae Z. Pearson of the Department of Development and Research, under whose direction most of the computations were carried out and who has checked through the various proofs.

Key to the Charts

The charts present various graphical representations of the function $W(c, X, n, N)$, equation 3. This function gives the probability, W, that the number of defectives in a lot of N is equal to *or less than* X, after a sample of n units has shown c defectives, *assuming that each of the possible values of* X *between* 0 *and* N *were equally likely a priori.*

Charts A: Separate pages refer to different values of n and N as labelled.

Ordinates, W; abscissas, X.

Solid curves, c; dotted curves $(c/n - X/N)$ expressed as per cent.

Charts B: Separate groups of curves refer to different values of W as labelled.

Ordinates, X/N; abscissas, n.

Separate sets of curves in each group refer to different values of c/n as labelled.

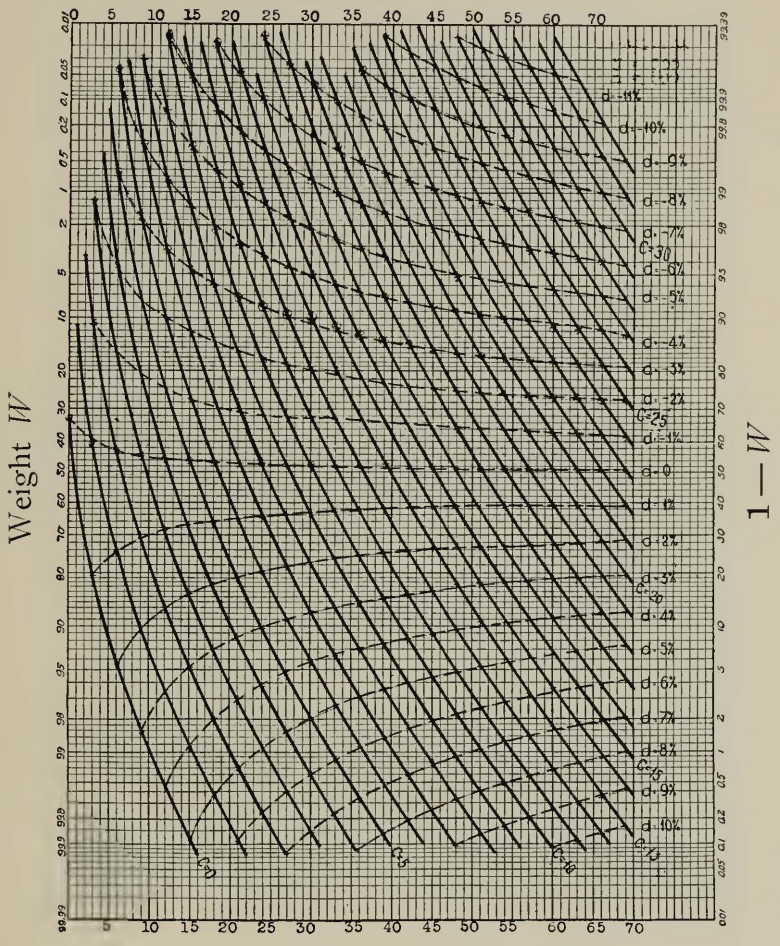
Individual curves are for different values of N.

Charts C: Separate pages refer to different values of W.

Ordinates, n; abscissas, n/N.

Separate curves for different values of c.

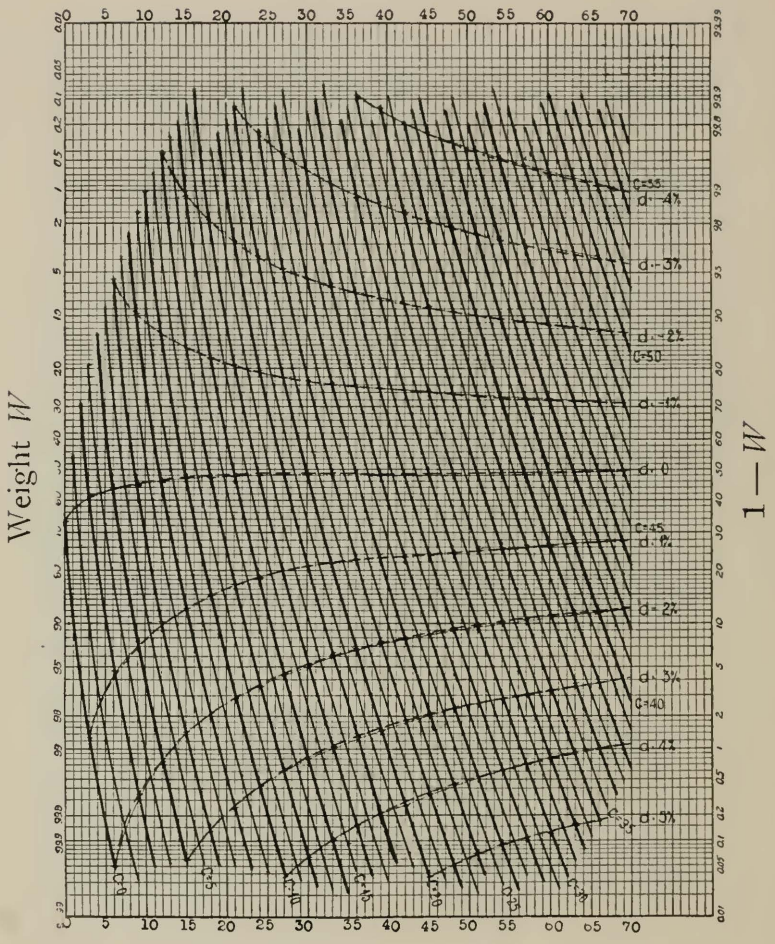
Cross-hatching indicates amount of dependence on X/N. For fuller explanation see pages 34–37 incl.

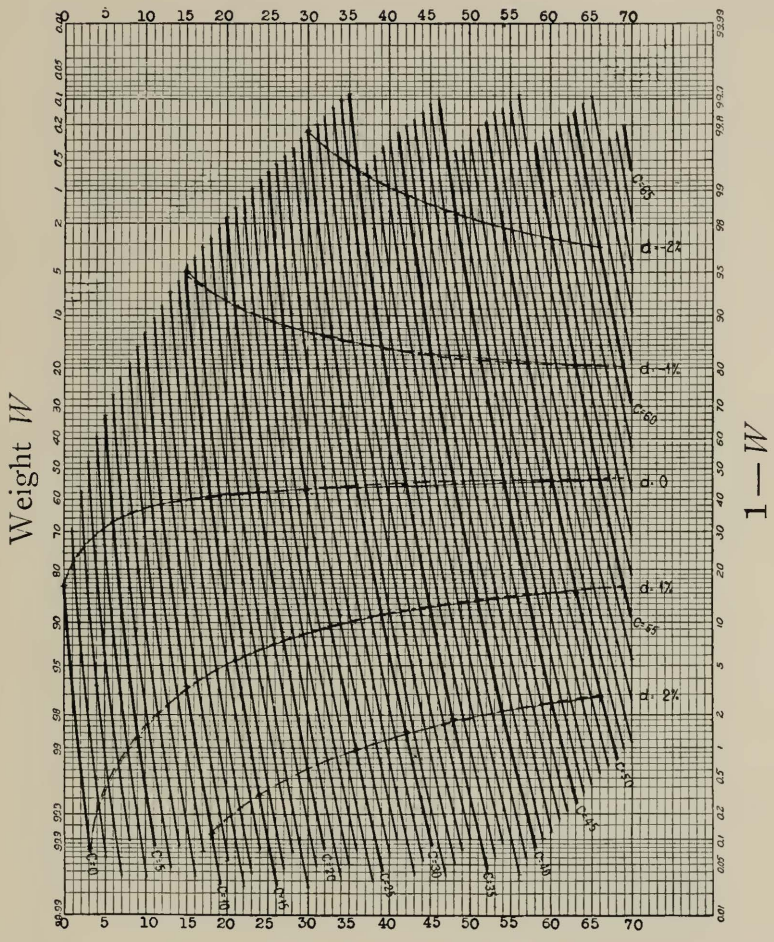


$X =$ Defectives in Universe

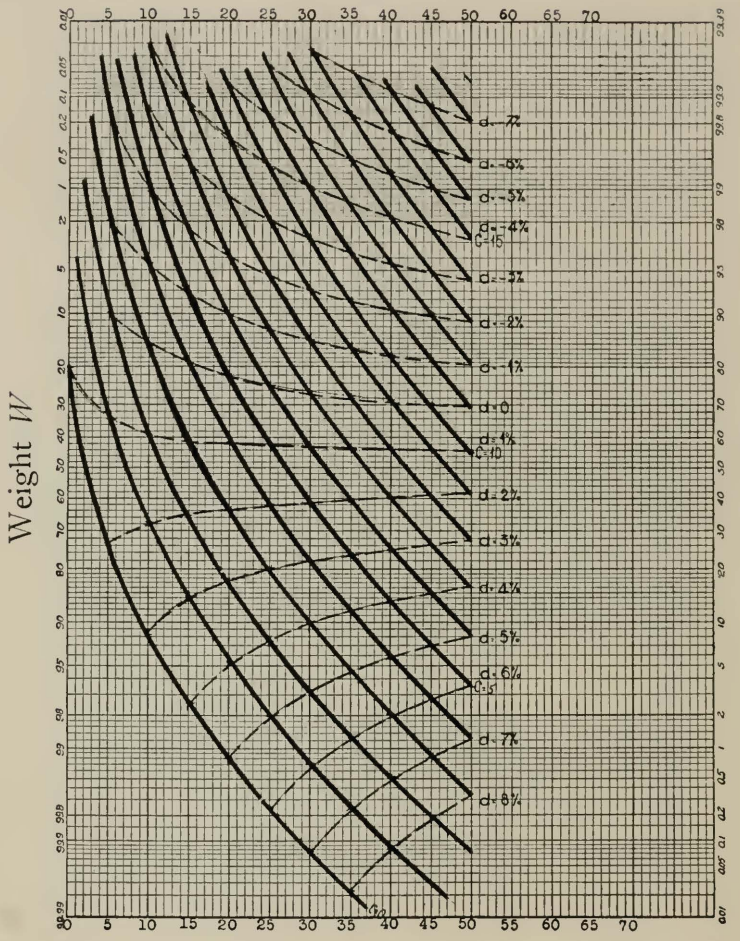
$N = 300,\ n = 100$

CHARTS A


Weight W
$1 - W$
X = Defectives in Universe
$N = 300,\ n = 199$
CHARTS A



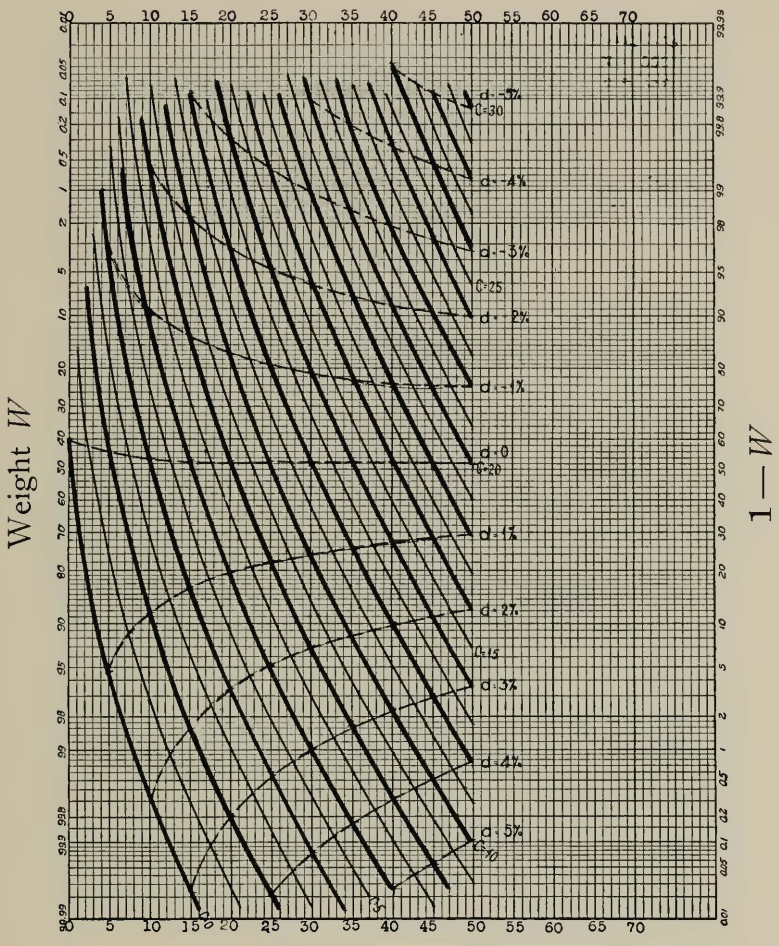
$X =$ Defectives in Universe
$N = 300,\ n = 249$
CHARTS A



$X =$ Defectives in Universe

$N = 500,\ n = 99$

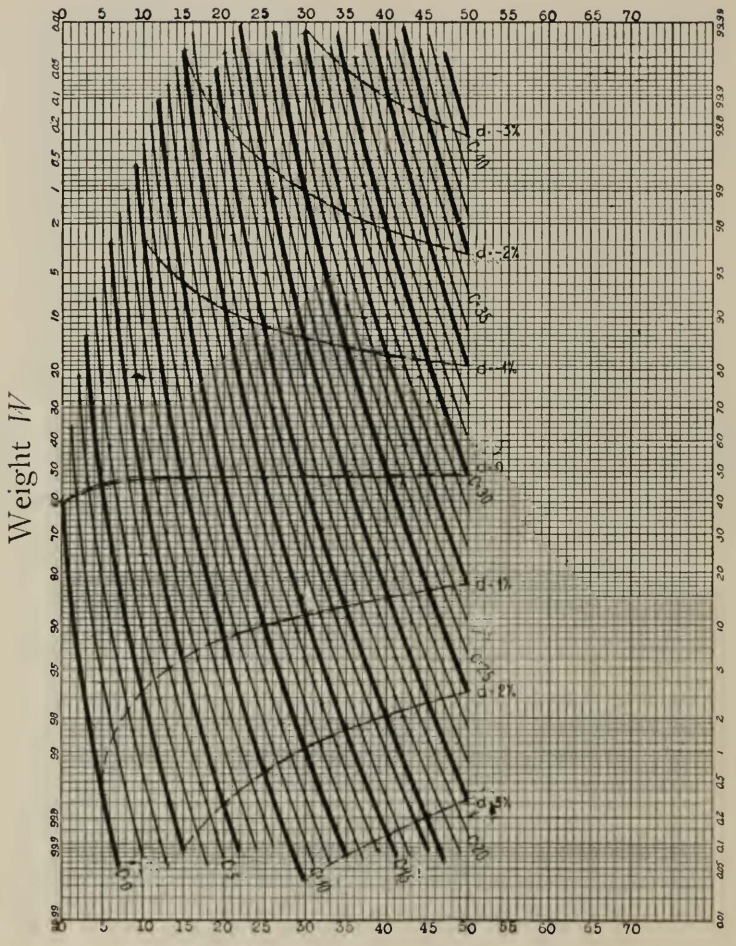
CHARTS A



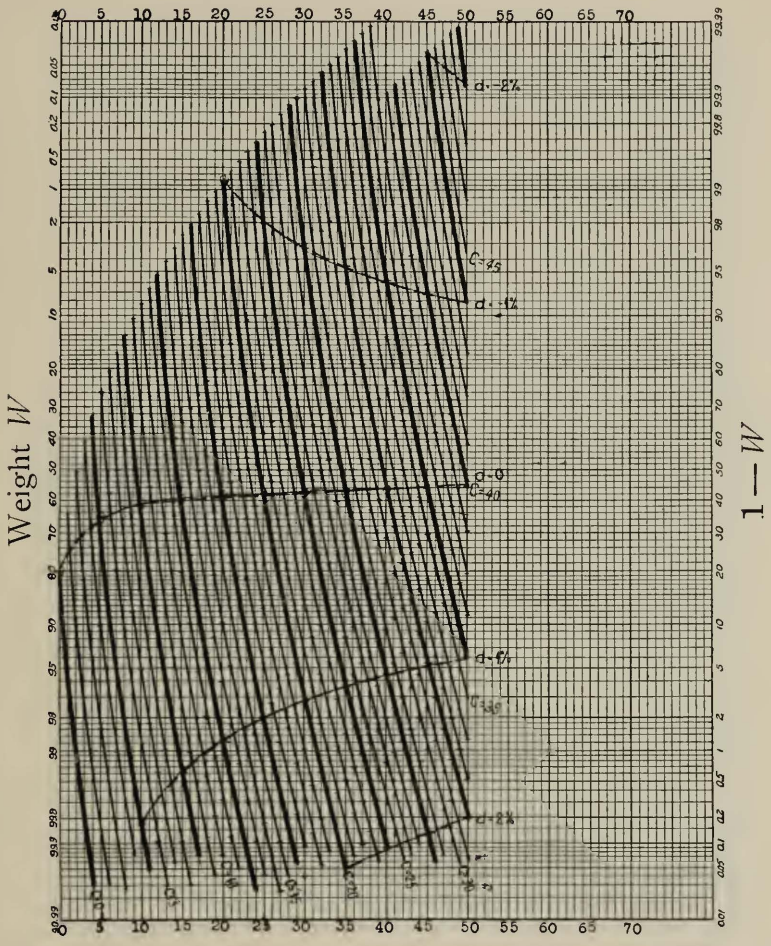
$X =$ Defectives in Universe

$N = 500,\ n = 199$

CHARTS A



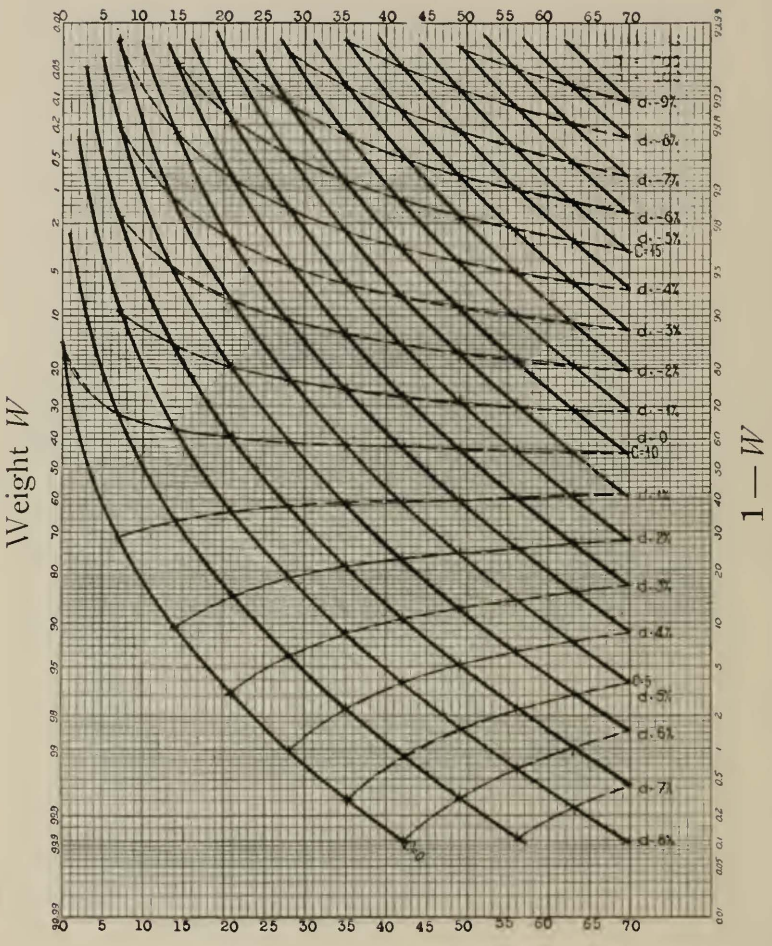
CHARTS A



$X =$ Defectives in Universe

$N = 500,\ n = 400$

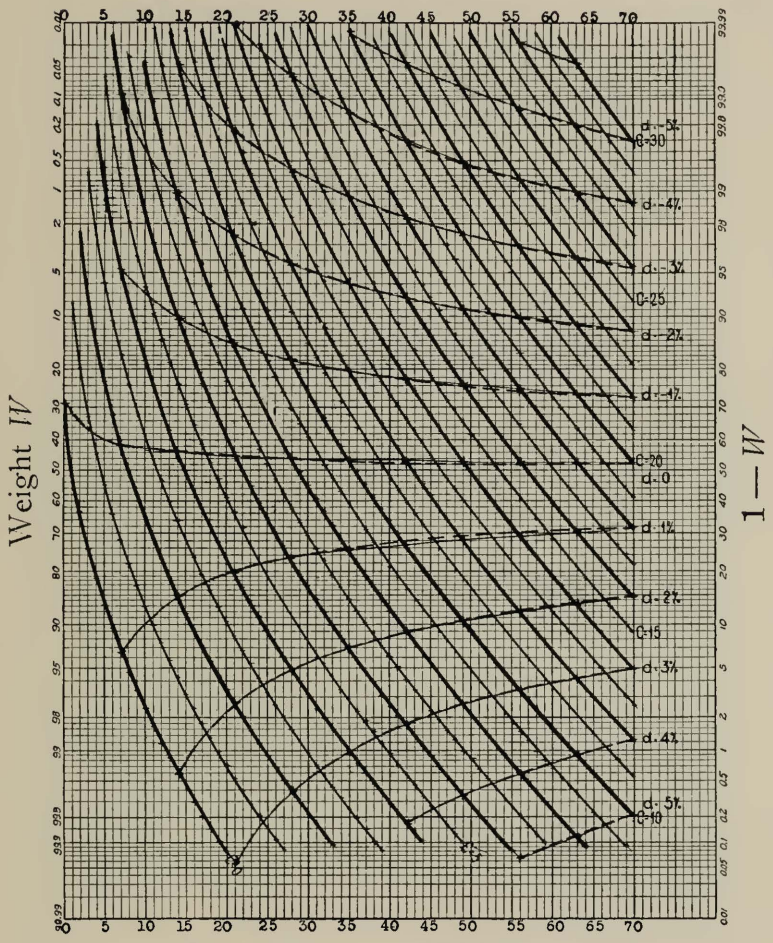
CHARTS A



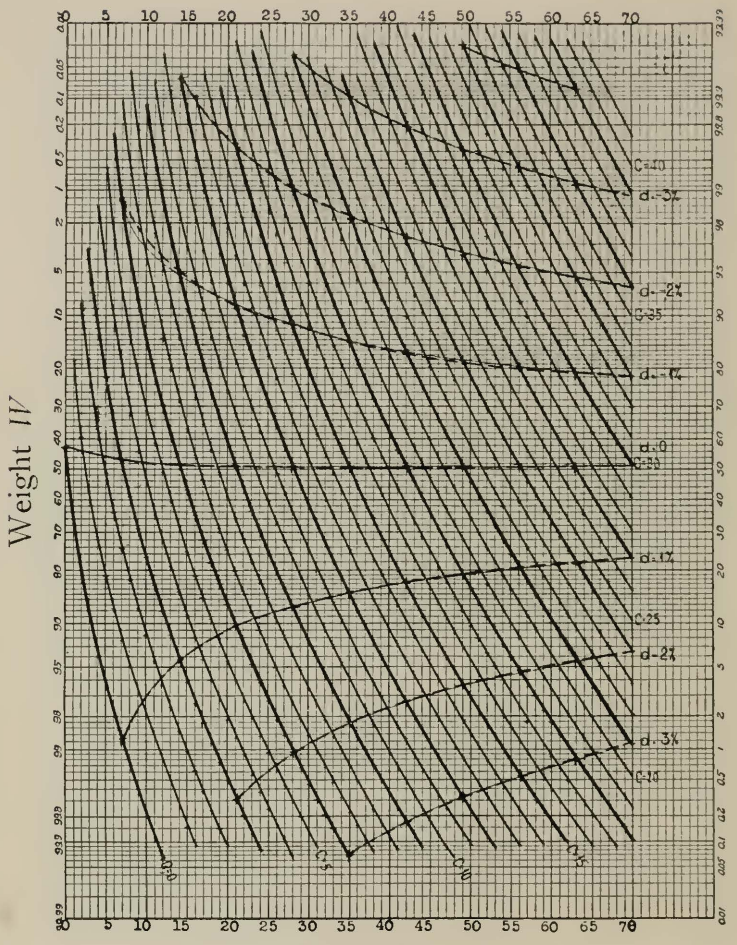
X = Defectives in Universe

$N = 700,\ n = 100$

CHARTS A



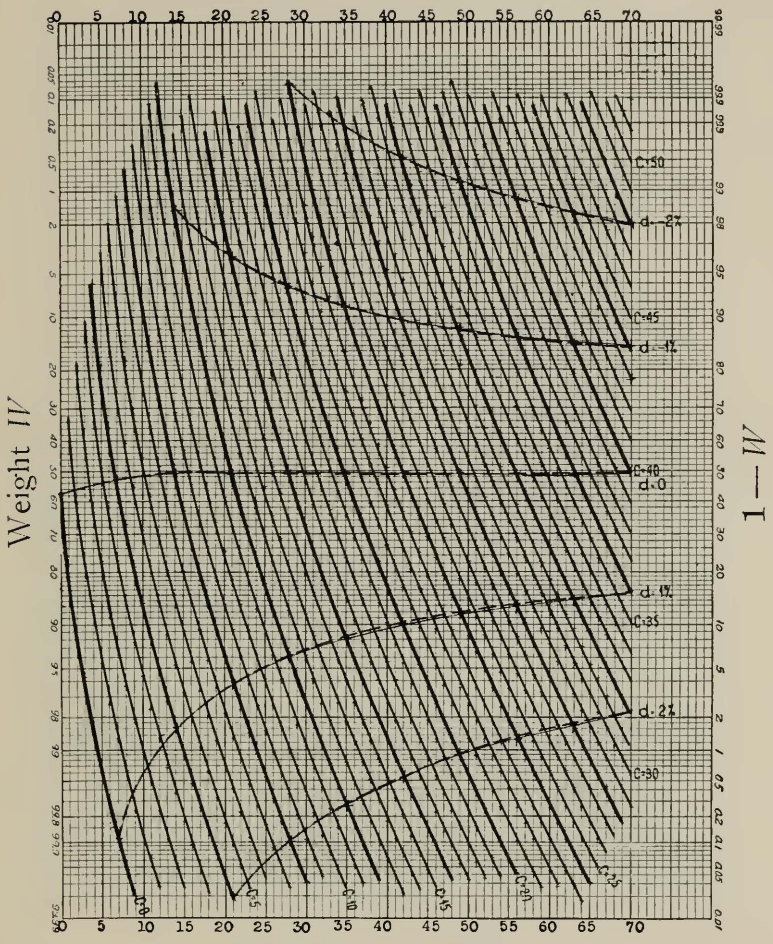
$X =$ Defectives in Universe
$N = 700,\ n = 200$
CHARTS A



$X =$ Defectives in Universe

$N = 700,\ n = 300$

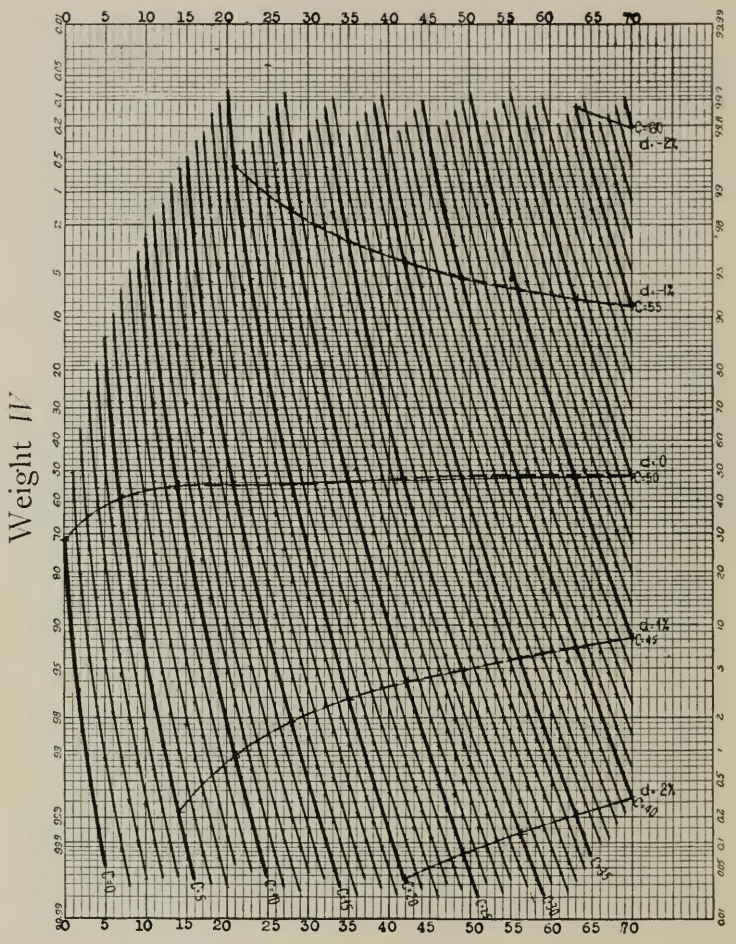
CHARTS A



$X =$ Defectives in Universe

$N = 700,\ n = 399$

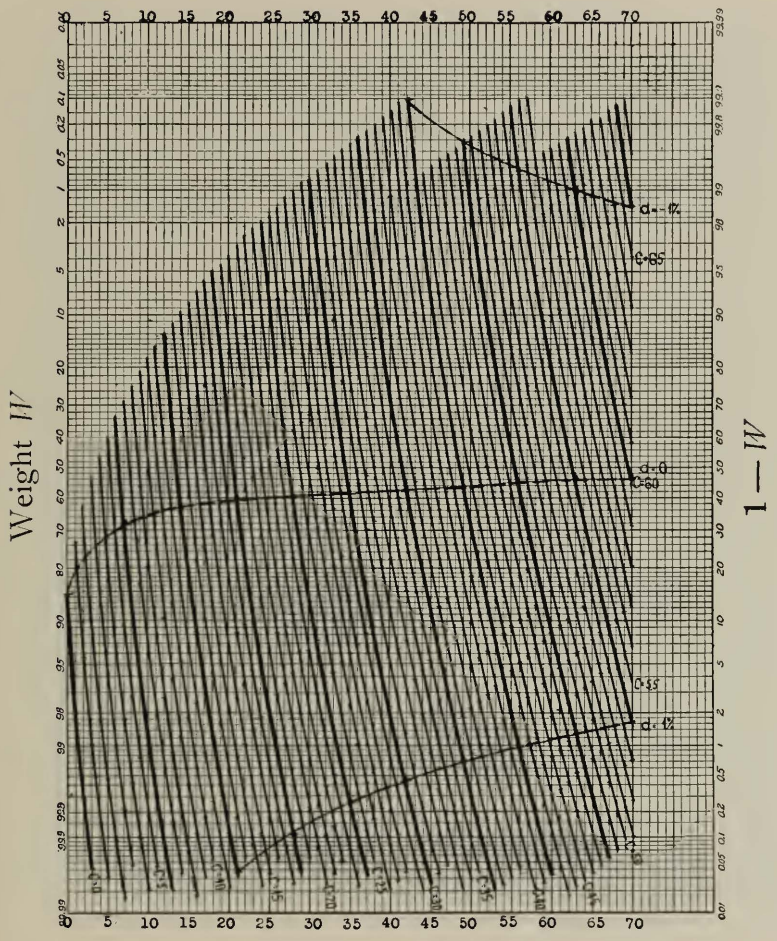
CHARTS A



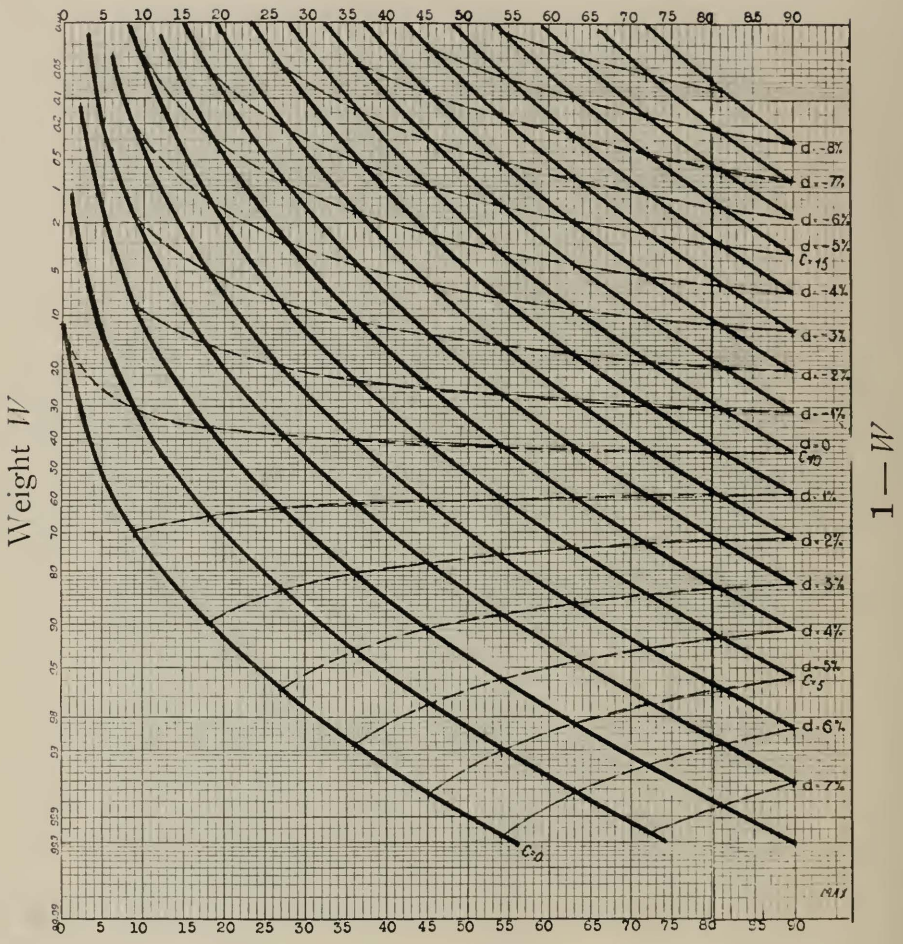
$X =$ Defectives in Universe

$N = 700,\ n = 499$

CHARTS A



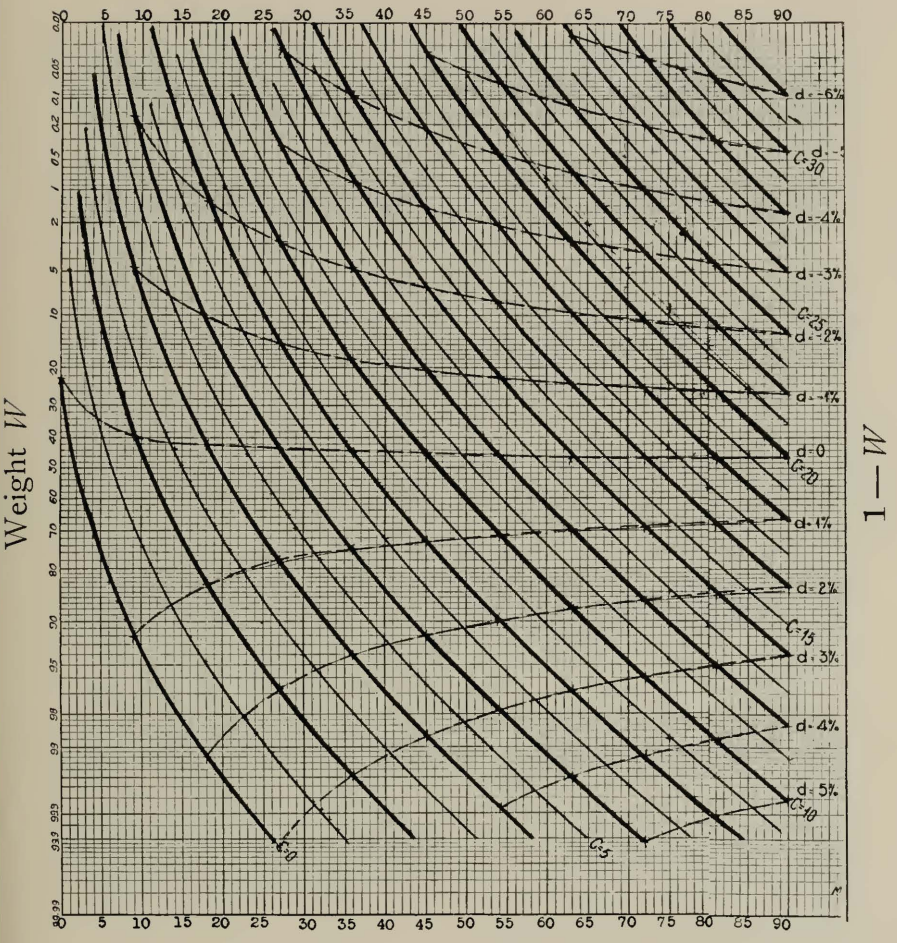
$X =$ Defectives in Universe
$N = 700,\ n = 599$
CHARTS A



$X =$ Defectives in Universe

$N = 900,\ n = 100$

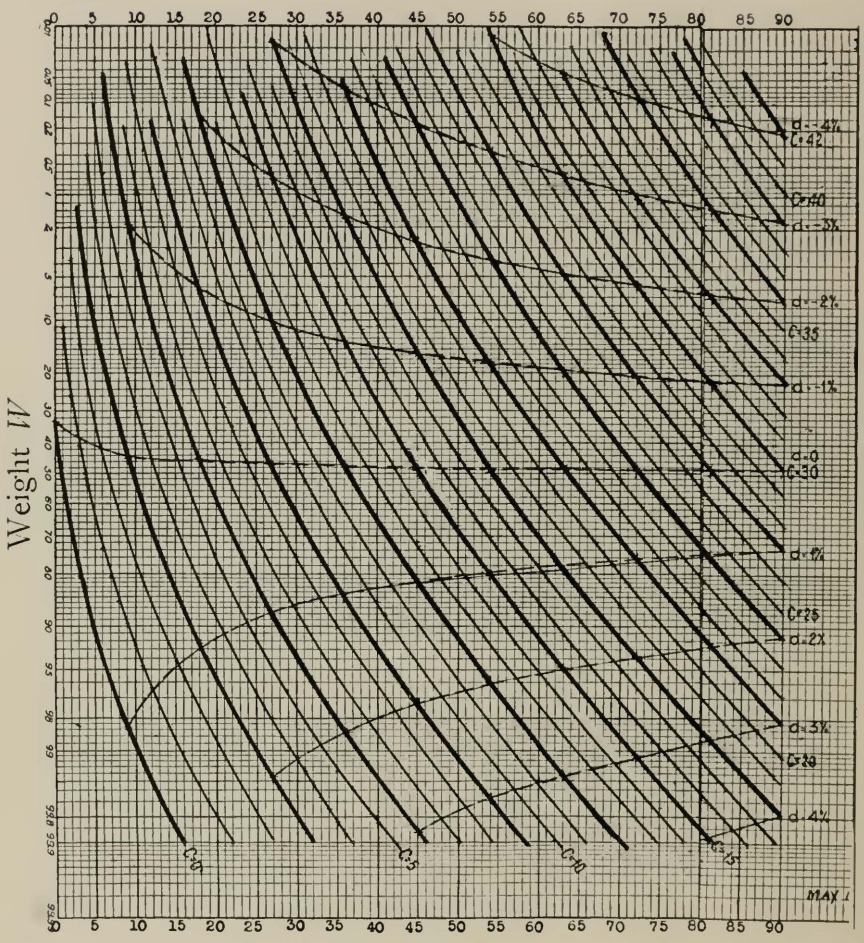
CHARTS A



$X =$ Defectives in Universe

$N = 900,\ n = 200$

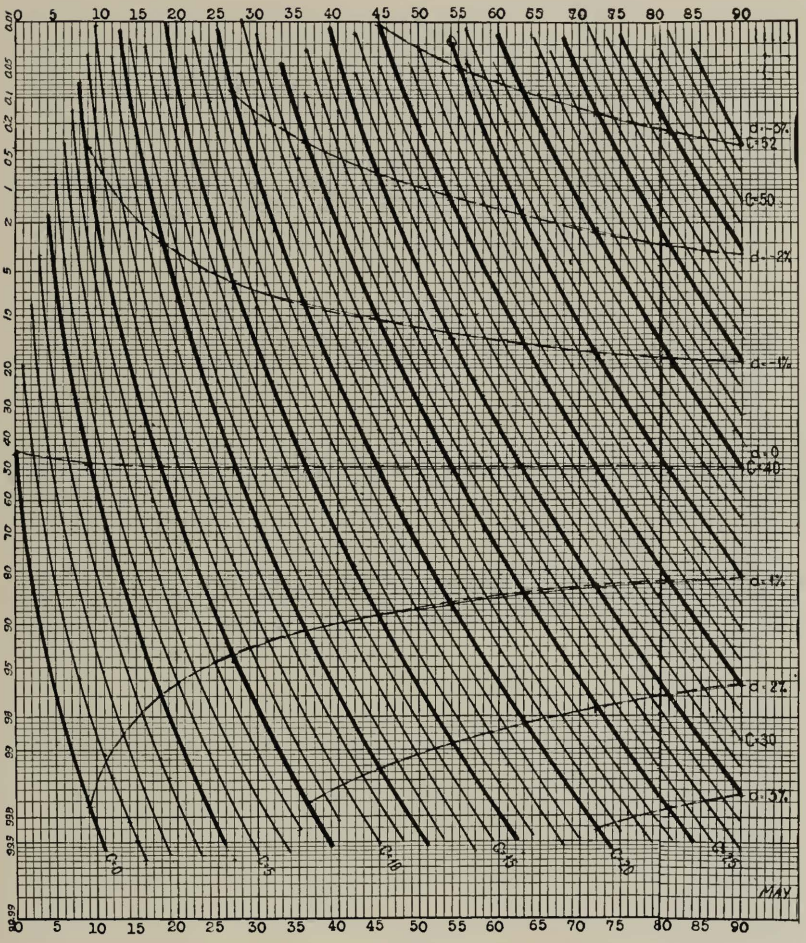
CHARTS A



$X =$ Defectives in Universe

$N = 900,\ n = 300$

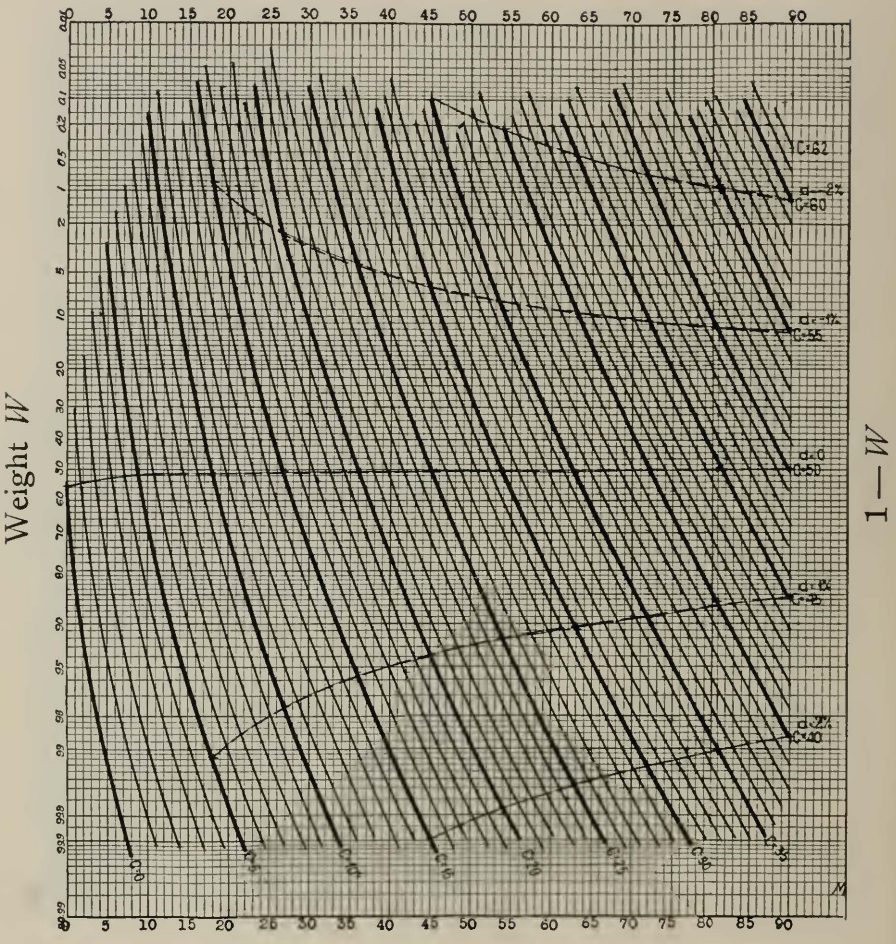
CHARTS A



$X =$ Defectives in Universe

$N = 900,\ n = 400$

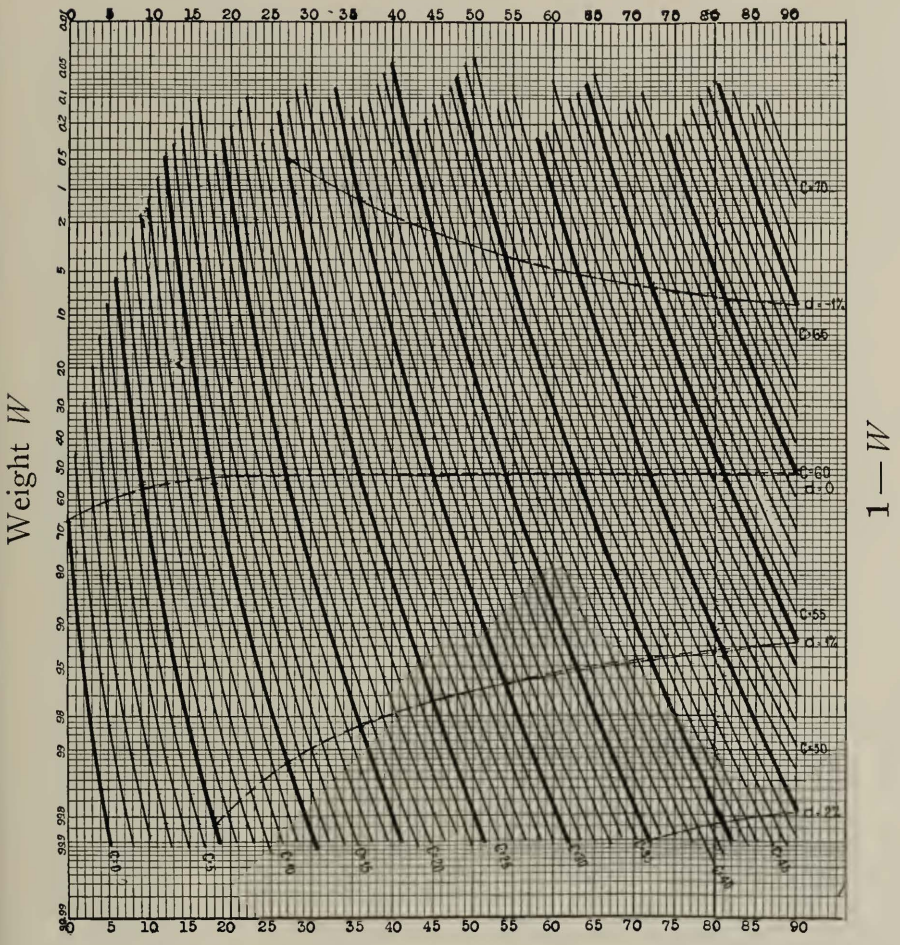
CHARTS A



$X =$ Defectives in Universe

$N = 900,\ n = 499$

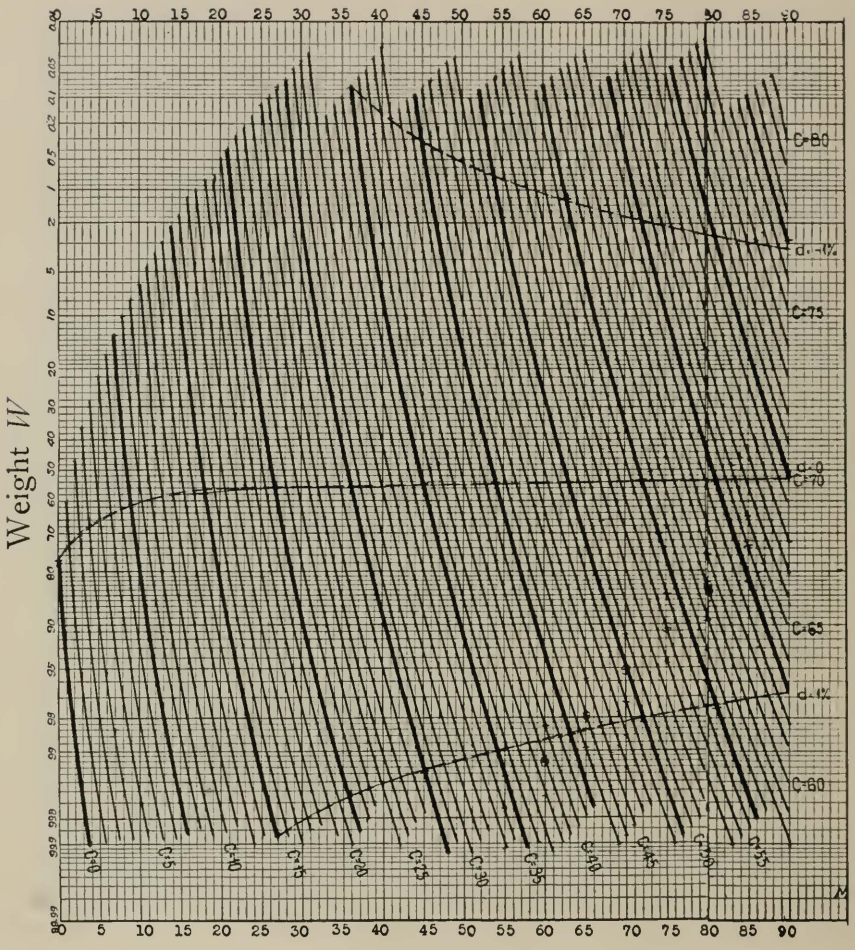
CHARTS A



$X =$ Defectives in Universe

$N = 900,\ n = 599$

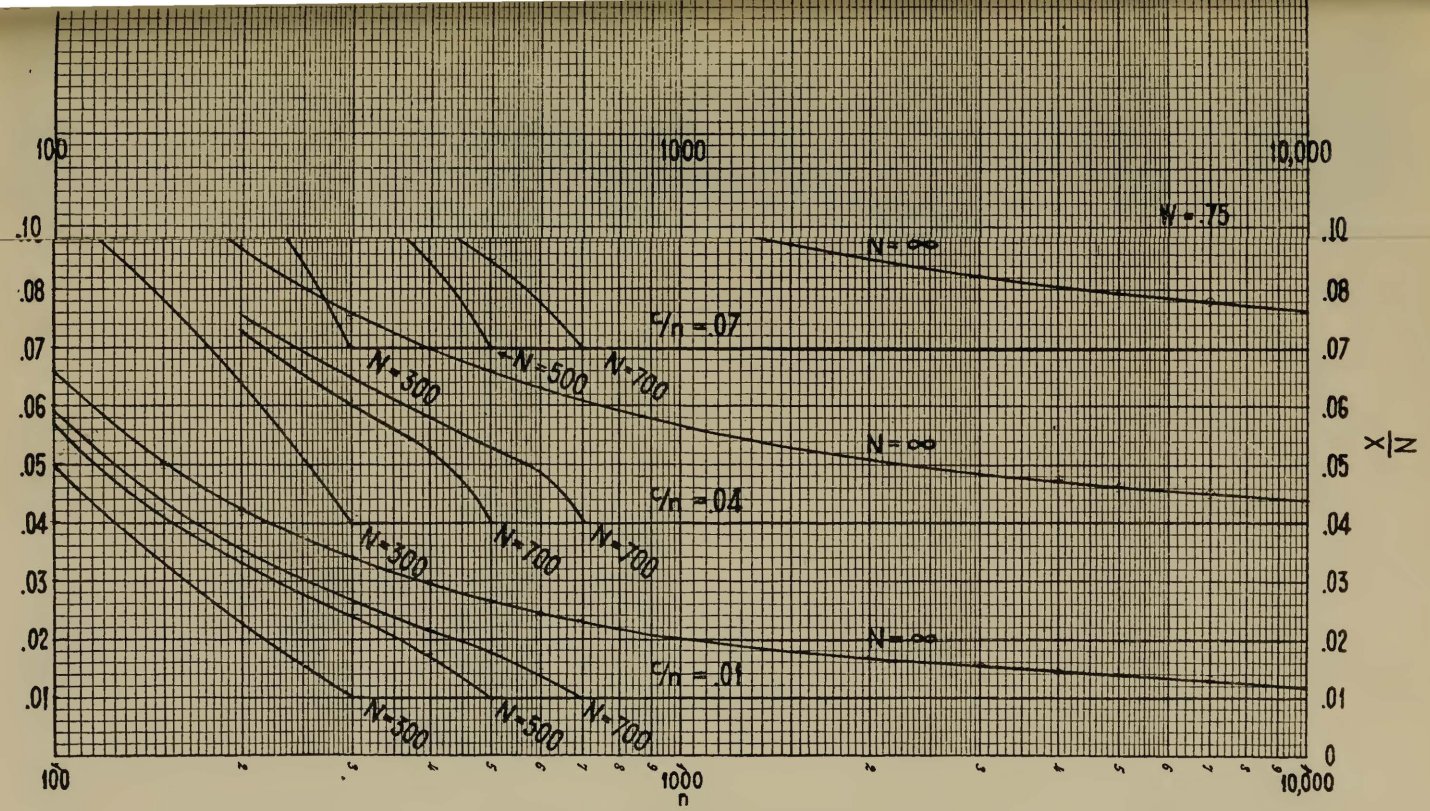
CHARTS A



$X =$ Defectives in Universe

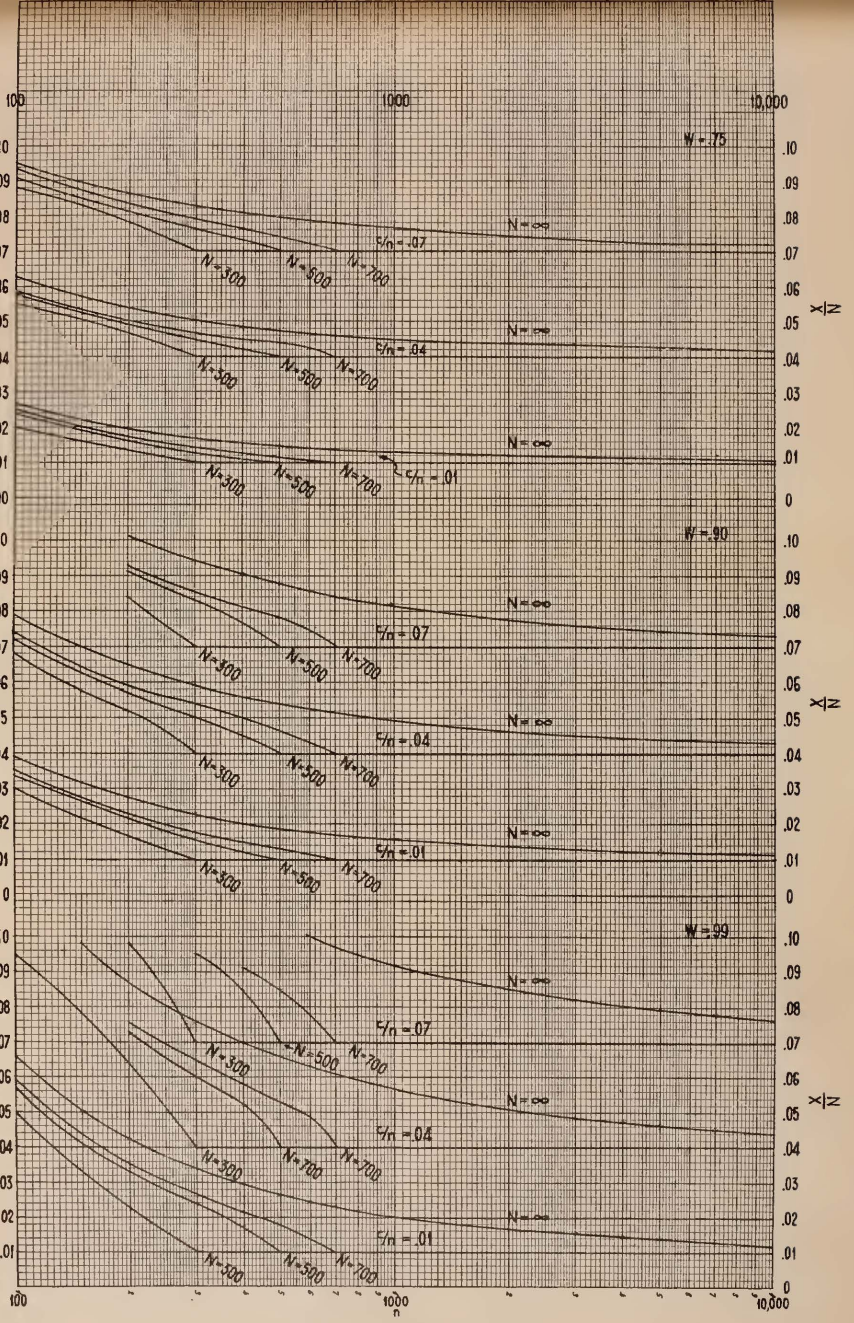
$N = 900,\ n = 699$

CHARTS A

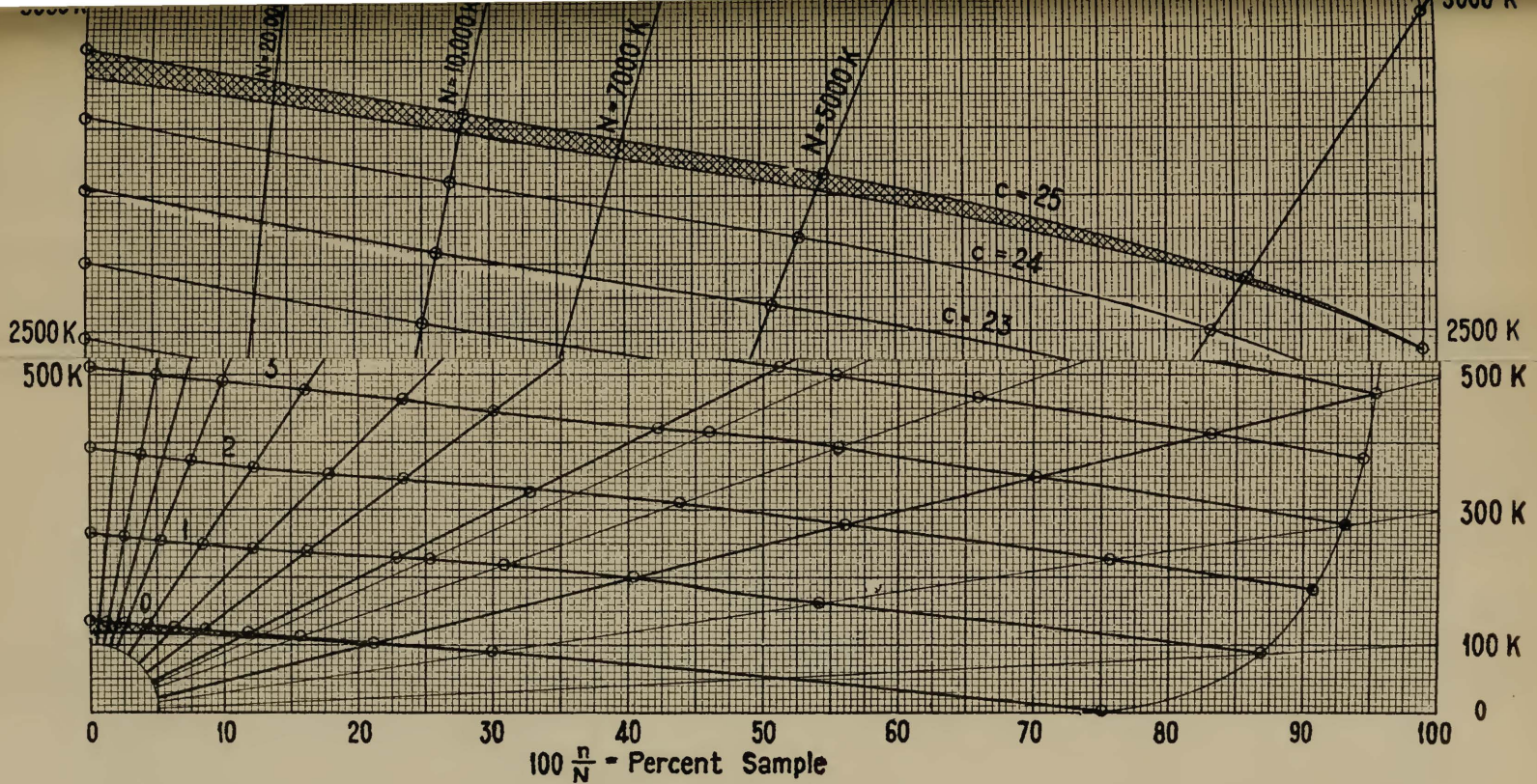


SAMPLING CHART B

$$\frac{\quad}{N} = \frac{\quad}{K} \qquad W = .75$$

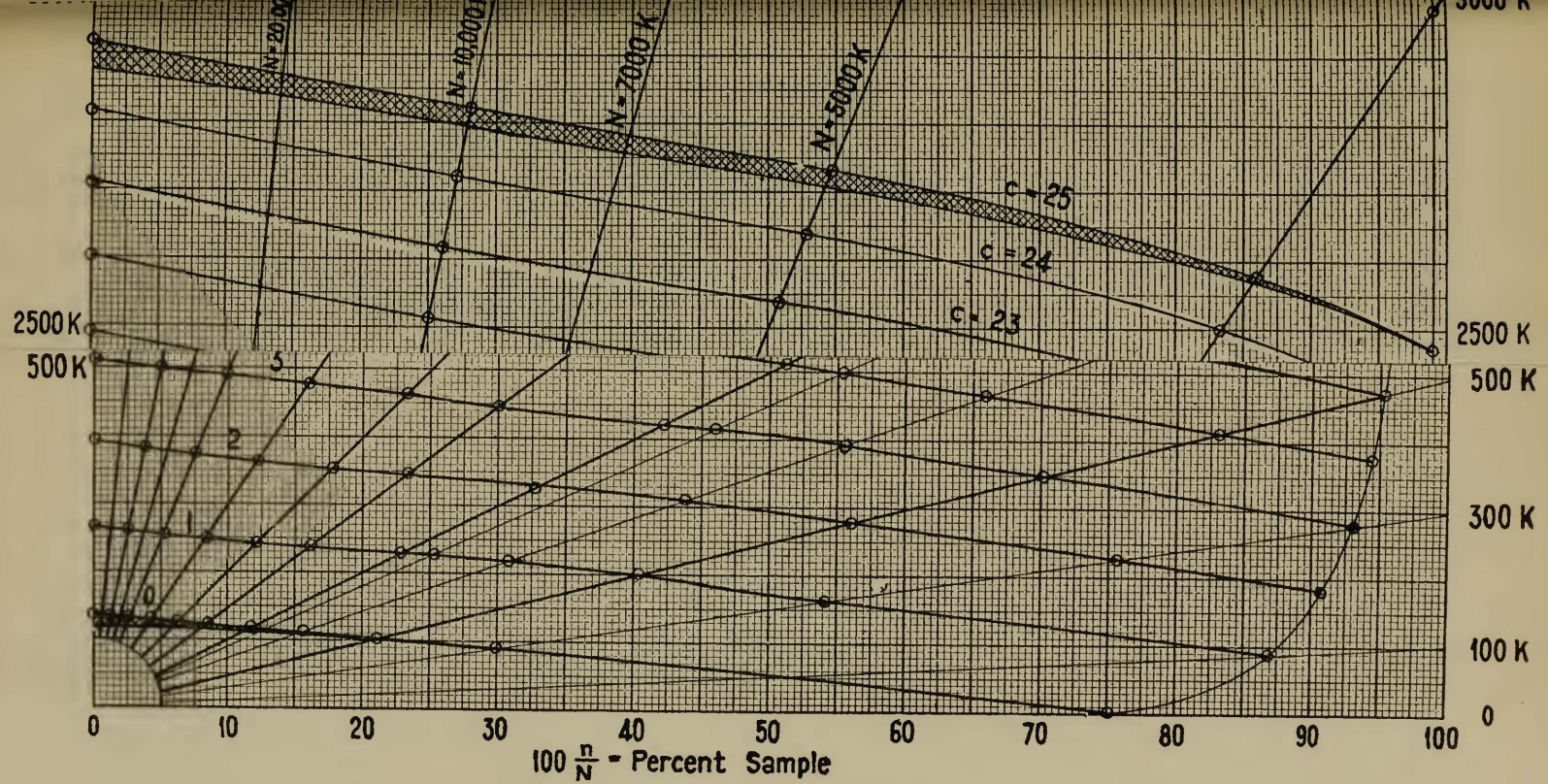


SAMPLING CHART B



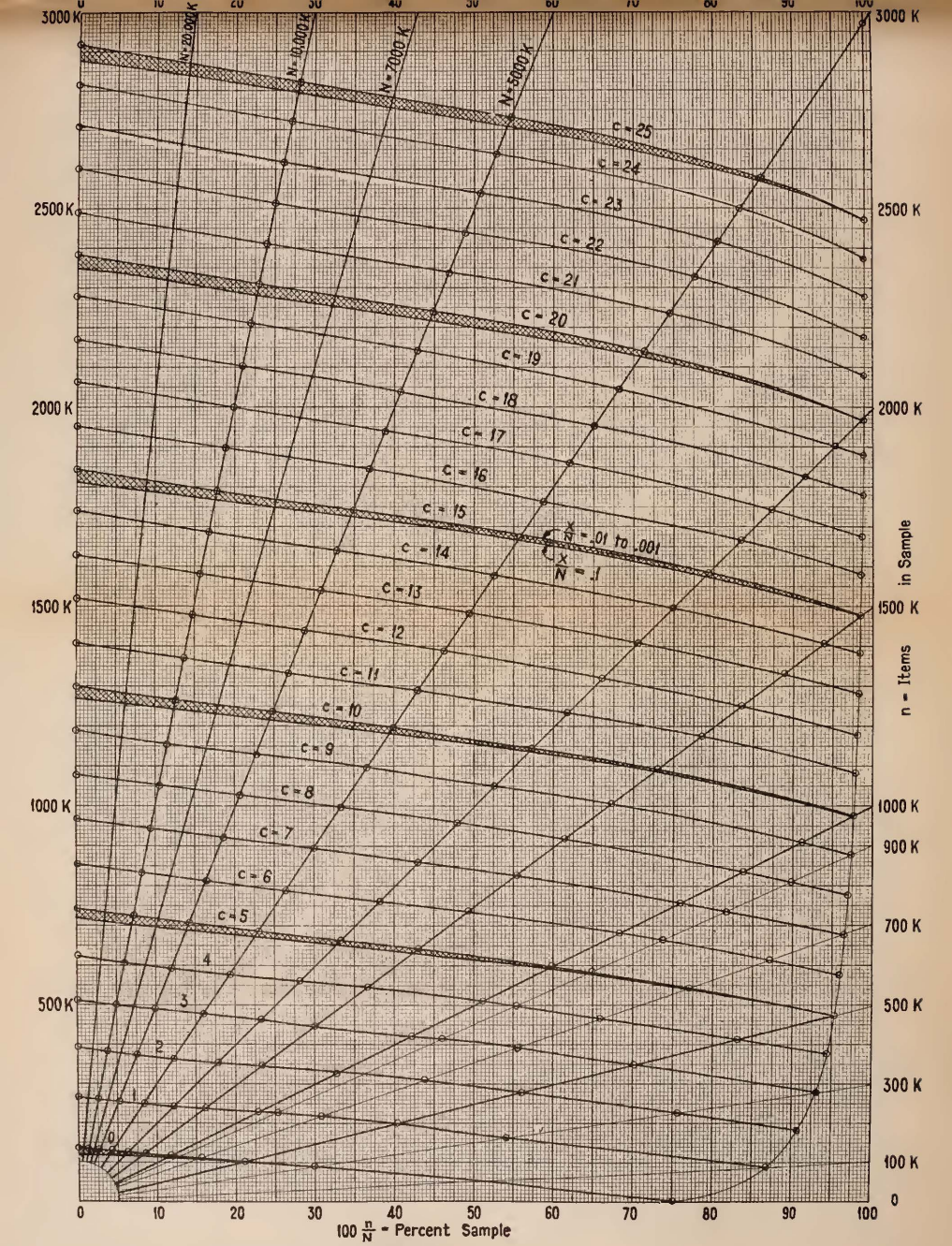
SAMPLING CHARTS C

$$\frac{X}{N} = \frac{.01}{K} \qquad W = .75$$



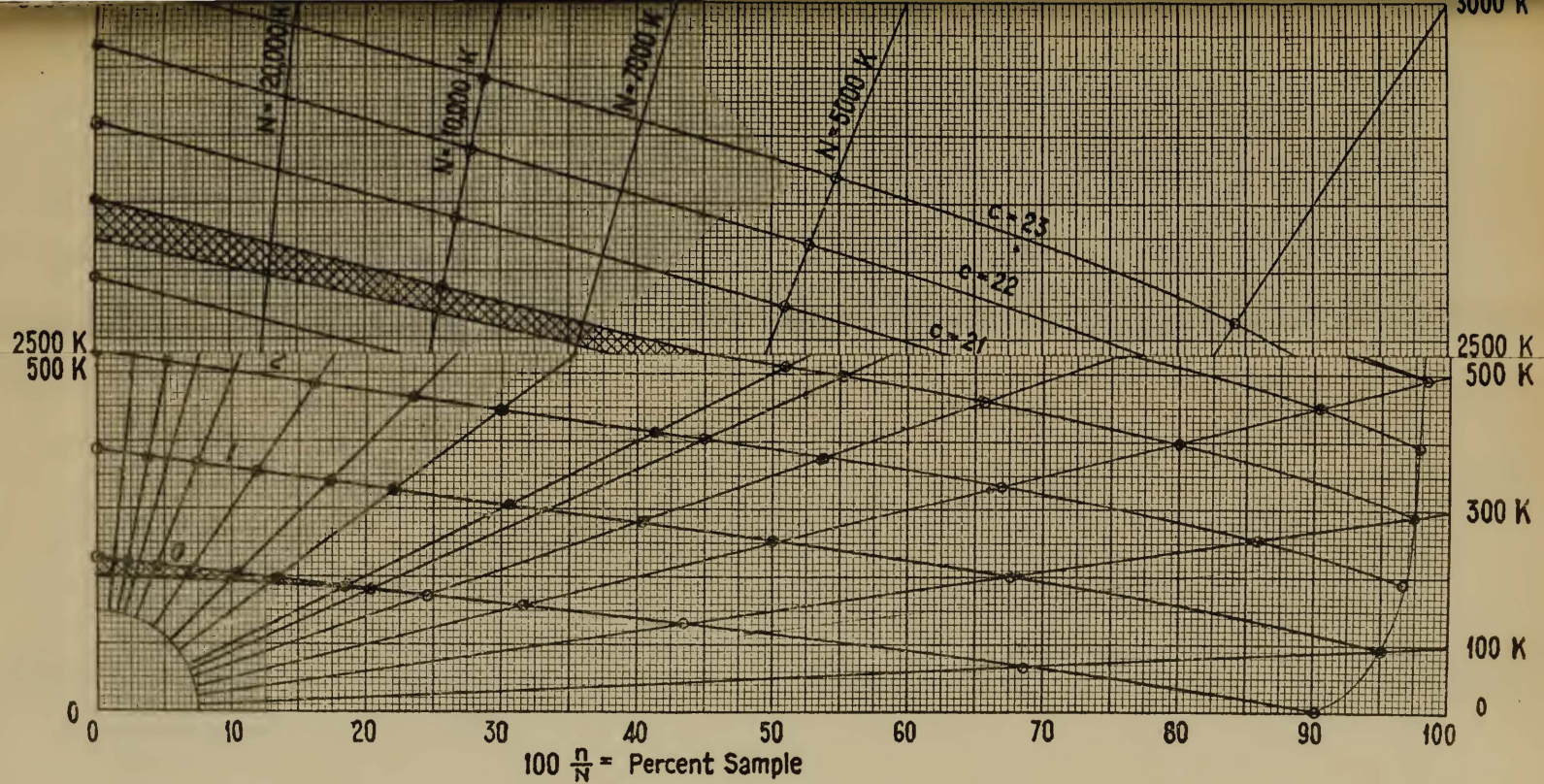
SAMPLING CHARTS C

$$\frac{X}{N} = \frac{.01}{K} \qquad W = .75$$



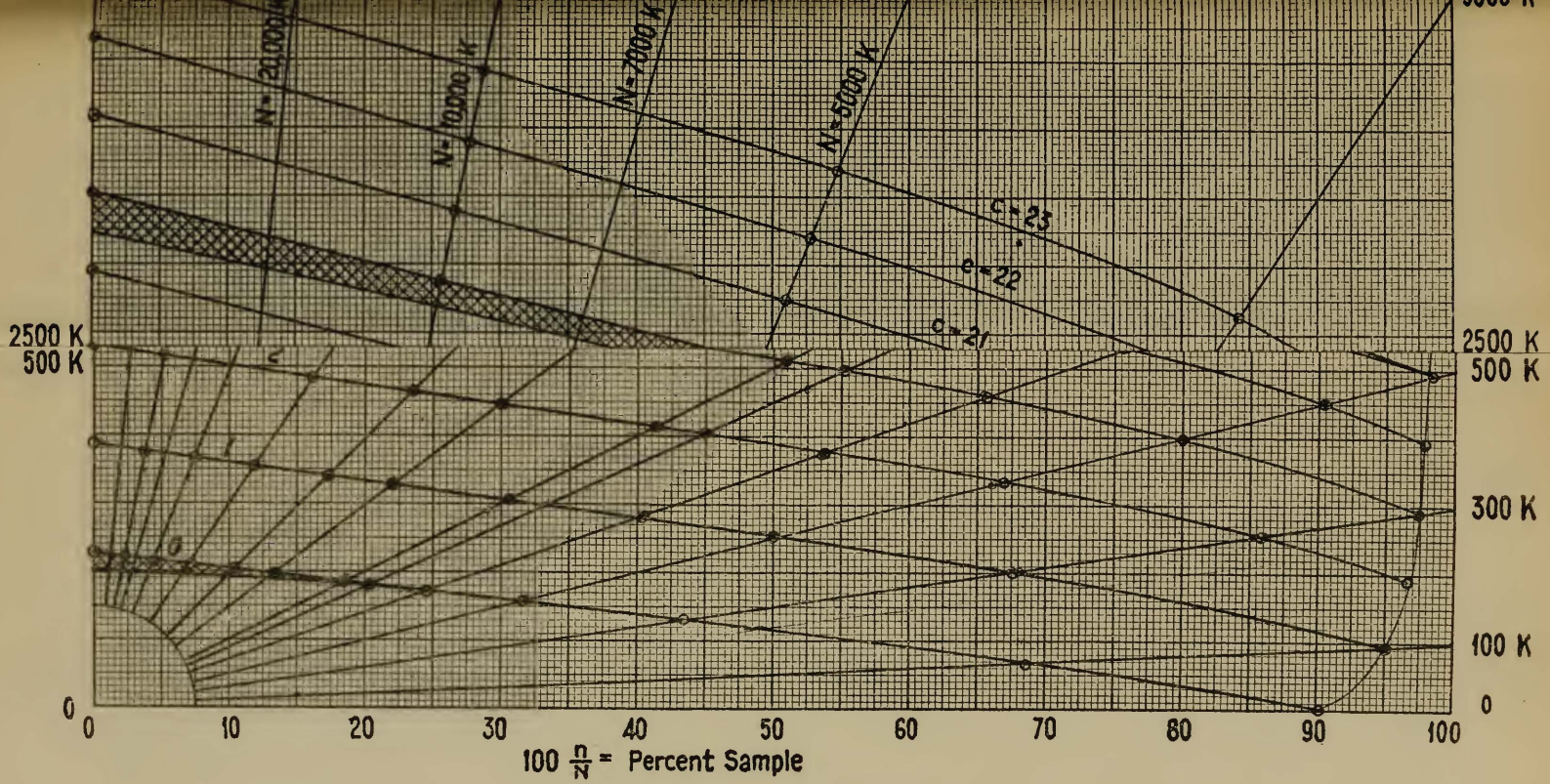
SAMPLING CHARTS C

$$\frac{X}{N} = \frac{.01}{K} \qquad W = .75$$



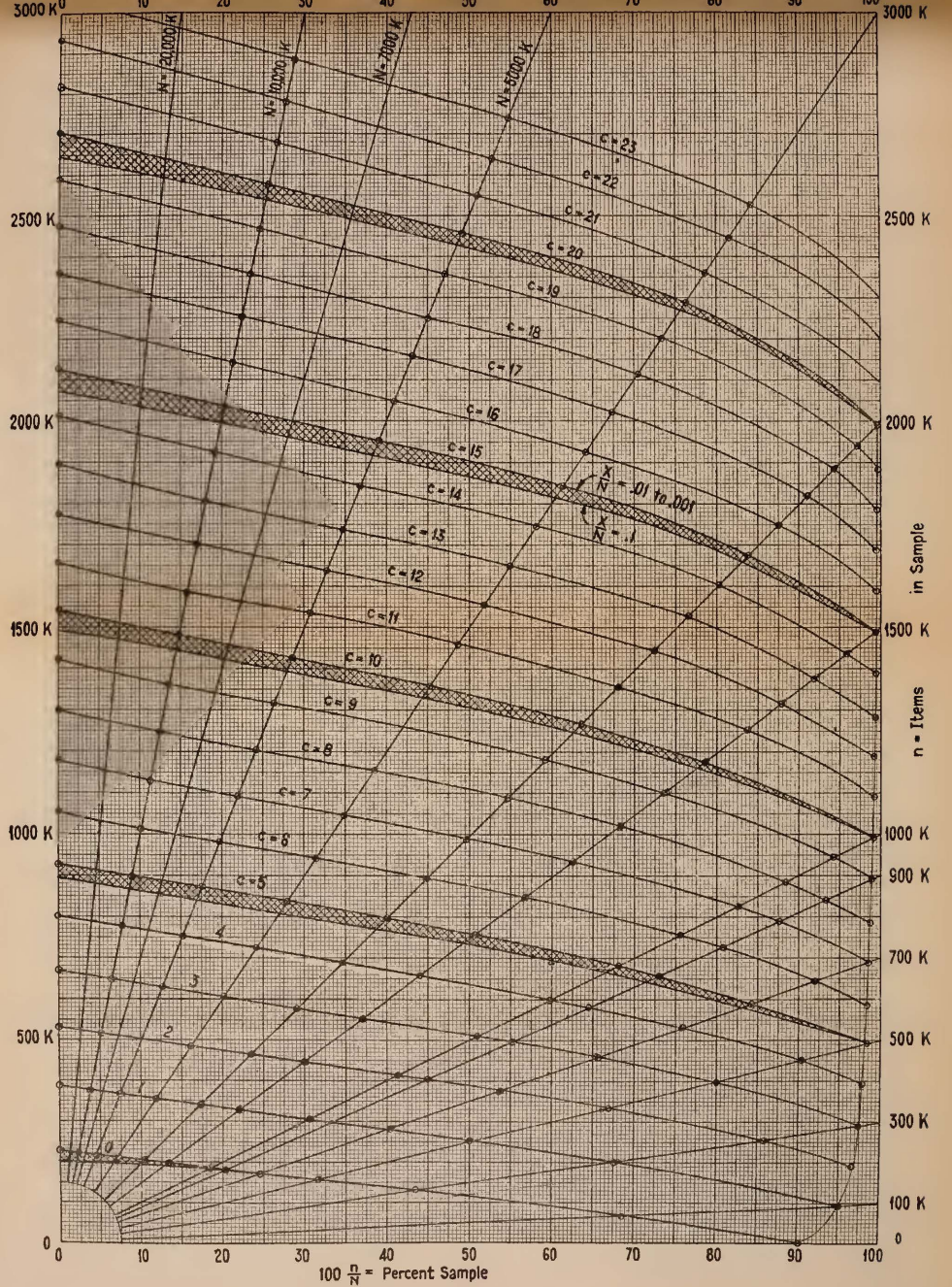
SAMPLING CHARTS C

$$\frac{X}{N} = \frac{.01}{K} \qquad W = .9$$



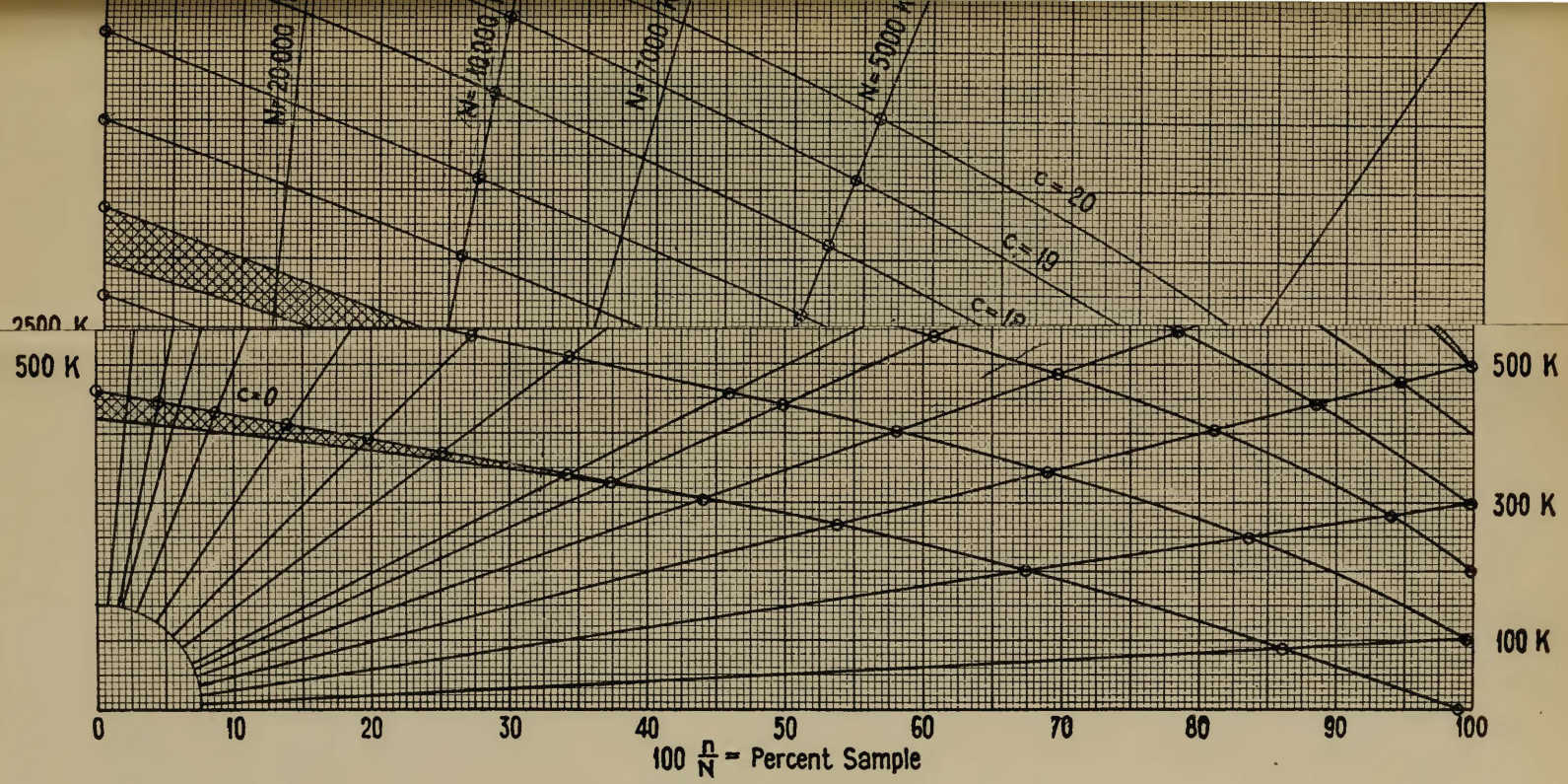
SAMPLING CHARTS C

$$\frac{X}{N} = \frac{.01}{K} \qquad W = .9$$



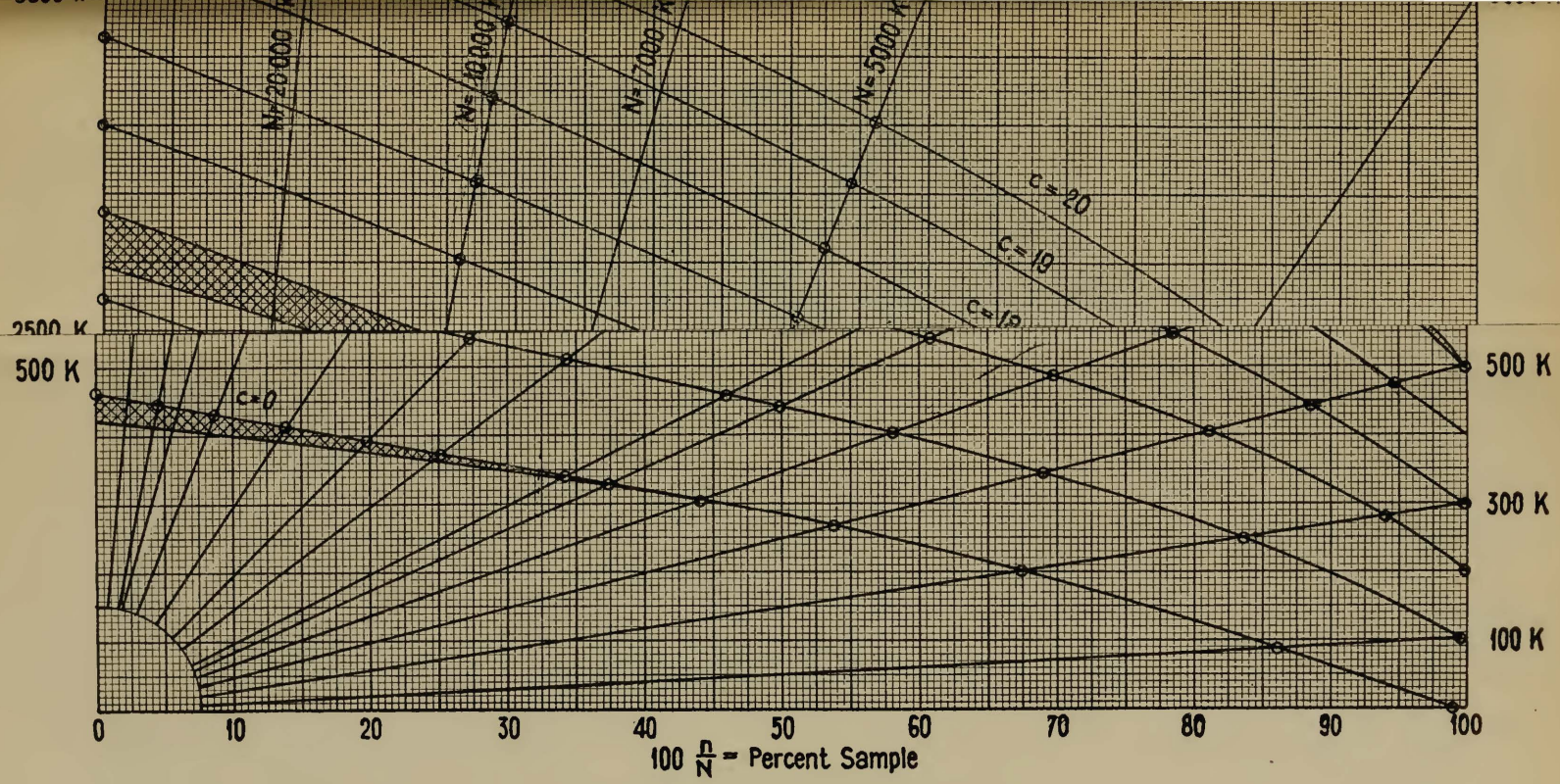
SAMPLING CHARTS C

$$\frac{X}{N} = \frac{.01}{K} \qquad W = .9$$



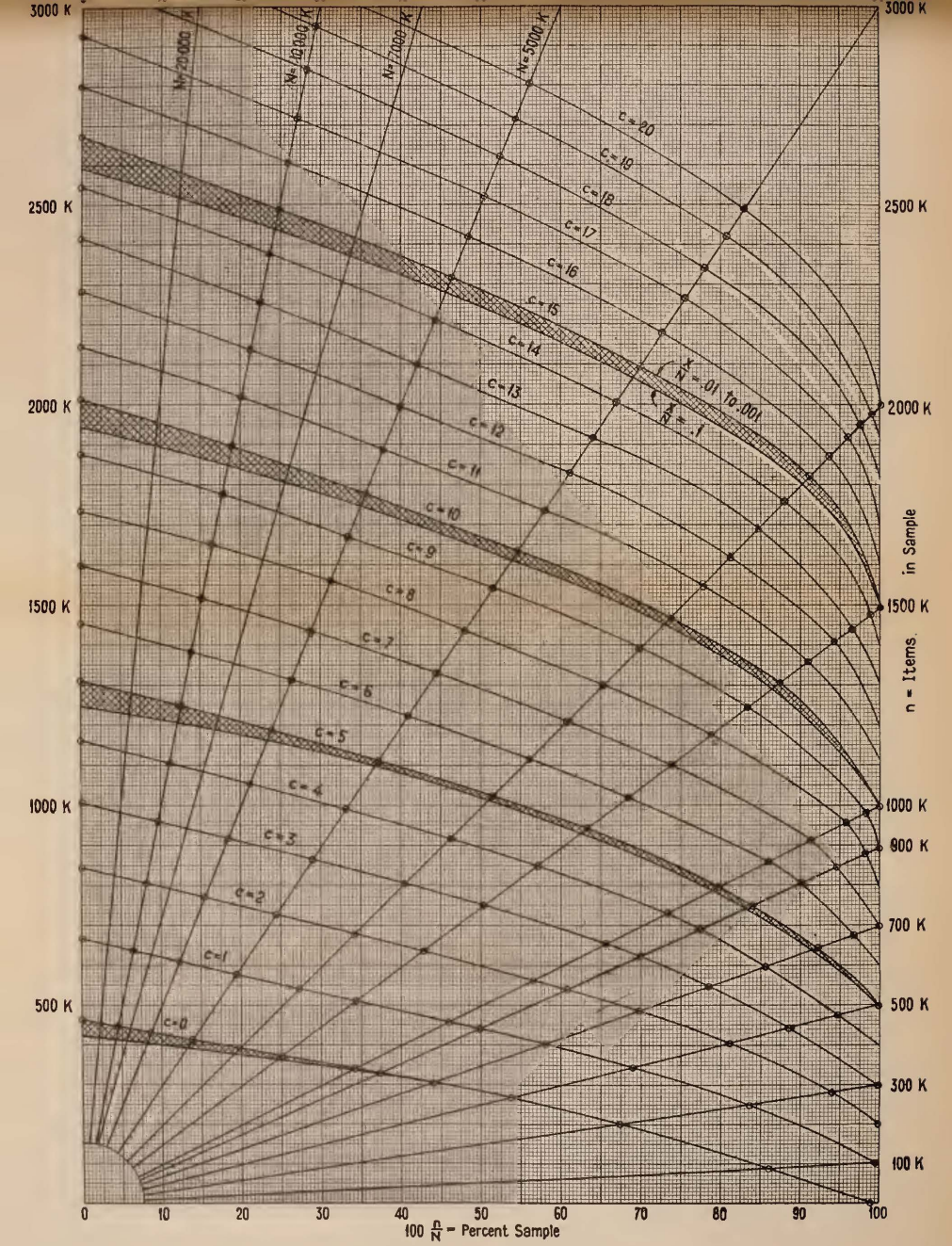
SAMPLING CHARTS C

$$\frac{X}{N} = \frac{.01}{K} \qquad W = .99$$



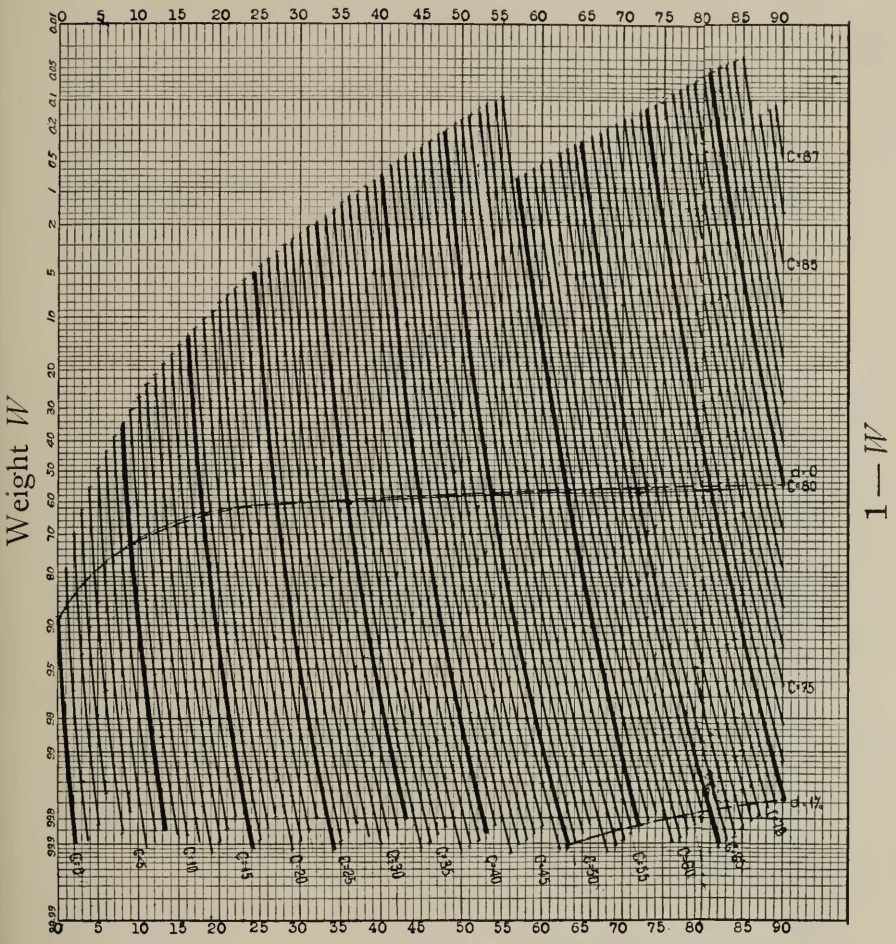
SAMPLING CHARTS C

$$\frac{X}{N} = \frac{.01}{K} \qquad W = .99$$



SAMPLING CHARTS C

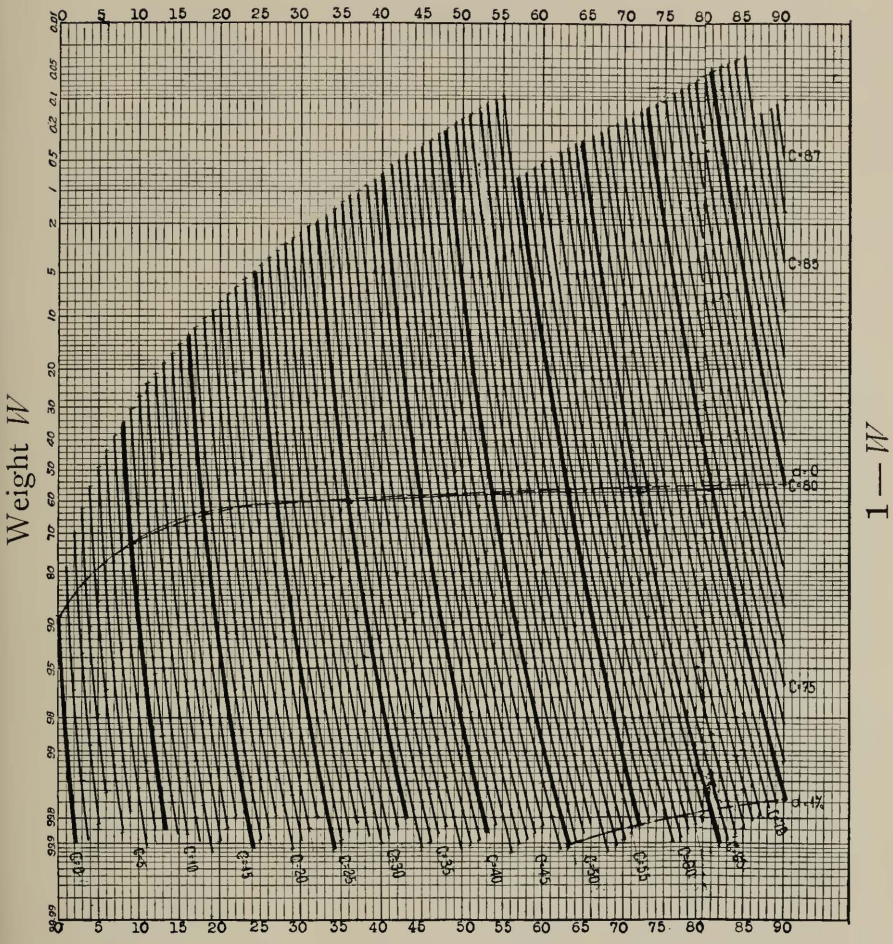
$$\frac{X}{N} = \frac{.01}{K} \qquad W = .99$$



$X =$ Defectives in Universe

$N = 900,\ n = 799$

CHARTS A



$X =$ Defectives in Universe

$N = 900,\ n = 799$

CHARTS A

Electrical Measurement of Communication Apparatus[1]

By W. J. SHACKELTON and J. G. FERGUSON

SYNOPSIS: This paper describes precision high-frequency measurements of a fundamental type, special emphasis being placed on the measuring circuits rather than on the types of apparatus measured. Standards of frequency, resistance, capacitance, and inductance are discussed briefly. Bridge measurements are described for the measurement of frequency, inductance, effective resistance, capacitance, dielectric loss, capacitance balance and inductance balance. Circuits for the measurement of other high-frequency characteristics such as attenuation, gain, and cross-talk are included.

INTRODUCTION

LONG DISTANCE electrical communication is now being effected by means of frequencies embracing the audible range and extending from there to the so-called short wave-lengths employed in radio transmission. According to the field of usefulness, this whole range has been subdivided into the audio, the carrier, and the radio ranges. From the viewpoint of the power engineer, all of the frequencies embraced in these ranges are high frequencies, but to the communication engineer, only those frequencies in the upper regions are considered high.

This paper discusses methods of measurement and measuring instruments adapted to the measurement of communication apparatus over this complete range. Most of the measuring apparatus described is designed particularly for use at audio and carrier frequencies. The measuring methods which are discussed are intended primarily for laboratory use in connection with the development and inspection of telephone apparatus prior to its application in the field.

Many of the transmission problems in the communication field involve the impedance characteristics of apparatus and circuits. In the manufacture of apparatus, impedance limits are used to a very great extent in inspection tests. Consequently, quantities of prime importance are those defining impedance characteristics; that is, inductance, capacitance and resistance at specified conditions, of course, such as temperature, frequency, and current or voltage. Other characteristics, of a less fundamental nature but nevertheless of considerable importance, are attenuation, gain, inductance and capaci-

[1] Presented at the Regional Meeting of District No. 1 of the A. I. E. E., Pittsfield, Mass., May 25–28, 1927.

tance balance, cross-talk, flutter and modulation. Since the three impedance components mentioned above, together with frequency, are probably of more general interest, this paper will be devoted largely to a discussion of their measurement, only brief reference being made to the methods used for the measurement of the latter group of characteristics.

As in all measurement work, standards representing the quantity are required, and these are of two classes, prime standards and secondary or working standards. In our case, the prime standards are resistance and frequency. From these we derive inductance and capacitance. Working standards are stable types of inductance coils, air and mica condensers, adjustable resistances, and for frequency, resonance type meters and highly stable oscillators.

Prime Standards

Frequency. The standard of frequency used is that described by Horton, Ricker and Marrison.[2]

Briefly, it comprises a special self-driven fork held at constant temperature and having all other conditions of operation so thoroughly controlled that a high degree of frequency stability is obtained. The exact frequency is measured by driving synchronously a phonic wheel for determining the number of cycles occurring in a given time interval. This time interval is usually a period of 24 hr. as measured by time signals received from Arlington. The average frequency of this fork is capable of being held constant and measured in this way with an accuracy of about 0.001 per cent.

The frequency of 100 cycles obtained from this fork is used to drive a 1000-cycle slave fork from which an equally constant 1000-cycle frequency is obtained. Having these frequencies, all other frequency measurements may be made with as high an accuracy as desired by direct comparison, using the cathode-ray tube as described in detail by Rasmussen.[3]

Resistance. Resistance standards specially designed for use with direct currents and having a very high degree of stability may be readily purchased or constructed and calibrations to a high degree of accuracy may be obtained from the Bureau of Standards. These resistance standards are not suitable for precision measurements at high frequencies, usually being wound on metal spools, and the value of the phase angle receiving only secondary consideration. It is

[2] Horton, Ricker and Marrison, "Frequency Measurement in Electrical Communication," *A. I. E. E. Transactions*, 1923.

[3] F. J. Rasmussen, "Frequency Measurements with the Cathode Ray Oscillograph," *A. I. E. E. Journal*, January, 1927.

necessary, therefore, to use resistance standards of special construction, depending upon the particular application to be made. In all cases, constancy of resistance with variations in atmospheric conditions, frequency and time is imperative. Generally as small a phase angle as possible is also highly desirable, although for some uses a suitable degree of constancy may be sufficient provided that the angle is known, and not large enough to affect appreciably the magnitude of the impedance of the resistance over the frequency range used.

To obtain the highest degree of stability of both resistance and phase angle, it has been found desirable to wind the wire on a spool made of a material not affected appreciably by atmospheric conditions, for example, phenol fiber, and to immerse the complete resistance in a sufficient amount of a suitable sealing compound to exclude all moisture. Resistances meeting all of the requirements outlined have been constructed as described in a recent paper by one of the authors.[4] Coils such as described there, having a resistance of approximately 1000 ohms, may be constructed to have an effective inductance of less than five microhenrys, and this inductance is practically independent of frequency up to at least 100 kc. Coils having lower values down to about 10 ohms can be made with equally small phase angles. Below this value of resistance, it is more difficult to hold a low phase angle.

Coils constructed as described may be considered to have so small a change in resistance with frequency that a calibration with direct current may be used without appreciable error for all frequencies at which they are used. Both the variation in resistance with frequency and the phase angle may be most readily measured by comparison with some simple type of resistance of such geometrical form that the phase angle may be readily computed. Satisfactory resistances for this purpose are short lengths of fine wire of definite shape, sputtered metal films on glass or other insulating material, and carbon in the form of rod or film.

Secondary Standards

Capacitance. The value of our capacitance standards is determined in terms of the prime standards of frequency and resistance. This determination may be made in several ways, the following bridge method being a simple and accurate one. The circuit, as shown in Fig. 1, consists of two equal resistance ratio arms, a resistance and capacitance in parallel in the third arm and a resistance and capacitance in series in the fourth arm. When this bridge is balanced at any particular frequency, the relations between the impedance arms

[4] W. J. Shackelton, "A Shielded A-C. Inductance Bridge," *A. I. E. E. Journal*, February, 1927.

of the bridge are such that the value of each capacitance may be determined in terms of the frequency and the two resistances.

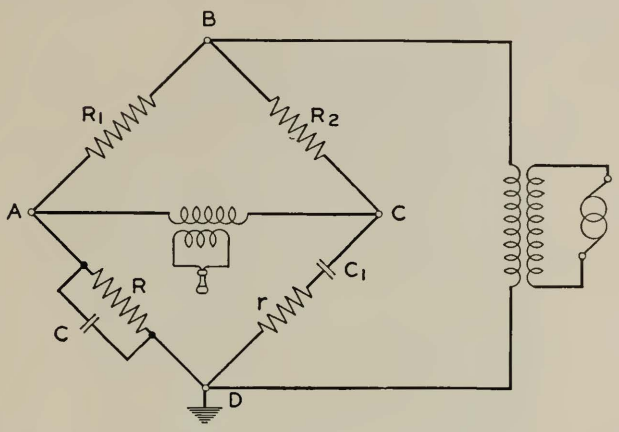


Fig. 1—Bridge circuit for measuring capacitance in terms of resistance and frequency

The requirements for a capacitance standard are high constancy with variations in frequency, time, voltage, and atmospheric conditions, and a small phase difference. Mica has been found to be the best solid dielectric, used either alone or impregnated with a high quality wax such as paraffin. If mica alone is used, the condenser must be sealed to prevent the entrance of moisture.

Good mica condensers can be obtained with a temperature coefficient below 0.005 per cent per deg. cent., and having a variation of less than 0.1 per cent over a frequency range from 500 cycles to 100 kc. Variations in capacitance with voltage are also negligible provided voltages below 100 volts are used. It has been our experience that the paraffin-impregnated condensers generally have a negative change of capacitance with temperature. This change is smaller than that of the unimpregnated type which has a positive change with temperature. The paraffin-impregnated condensers, however, usually change more with time than the unimpregnated condensers.

Air condensers may be used as standards in small sizes. For the larger values, the air condensers become large and cumbersome and are not as stable as the mica condensers. Even in the smaller sizes, very special precautions must be taken to obtain air condensers which have appreciably smaller phase differences than the mica condensers, which may be made with phase differences considerably less than one minute.

Inductance. Requirements for inductance standards are high con-

stancy with variations in time, current or saturation, atmospheric conditions, and frequency. It is also desirable that they be made with a small external field. Otherwise, very great care must be taken to avoid errors due to this cause.

In order to obtain stability with variations in saturation, it is usual to make inductance standards with air cores. This requires standards of large physical size if a time constant as large as the average iron core coil is desirable. This large size results in large capacitance distributed in the coil itself and from the coil to ground. These capacitances cause large variations in inductance with frequency and with the position of the coil with respect to ground. On account of this difficulty with air core coils, permalloy [5] as core material has been used with considerable success as described by one of the authors. [4]

The calibration of these inductance standards may be made by comparison with any two of the quantities, capacitance, resistance and frequency. Comparison with frequency and resistance may be made in a bridge circuit exactly similar to the one used for capacitance determination, substituting inductances for capacitances. A comparison with frequency and capacitance may be made by means of a resonant method, and comparison with capacitance and resistance may be made by means of the Owen bridge.[6] The resonant method is used generally except for those cases requiring large capacitance, in which cases the Owen bridge is used.

Frequency. As a secondary standard of frequency for use with th cathode ray tube, where practically only one standard frequency is required, a special 1000-cycle oscillator is used, designed particularly for high stability of frequency with ordinary variations in external conditions. This oscillator is shown in Fig. 2. It allows the use of a cathode ray tube for frequency measurements with a high degree of accuracy under conditions where the prime standard of frequency is not accessible.

Where a portable frequency standard is desirable, for instance, as a means of shop frequency checks, a resonance type of meter is used. This is shown in Fig. 3. It is essentially a resonance bridge circuit consisting of two equal resistance ratio arms, a third arm containing a resonant circuit, and a variable resistance as the fourth arm. The capacitance and resistance are variable over wide ranges by means of decade switches, and the capacitance is capable of fine variations by the use of a form of precision variable air condenser having provision for fine control. There are four air-core inductance coils which give,

[5] H. D. Arnold and G. W. Elmen, *Franklin Institute Journal*, Vol. 195, 1923.

[6] D. Owen, "A Bridge for the Measurement of Self-Inductance," *Proceedings of the Physical Society of London*, October, 1914.

in conjunction with the variable capacitance, a frequency range of about 100 cycles to 150 kc.

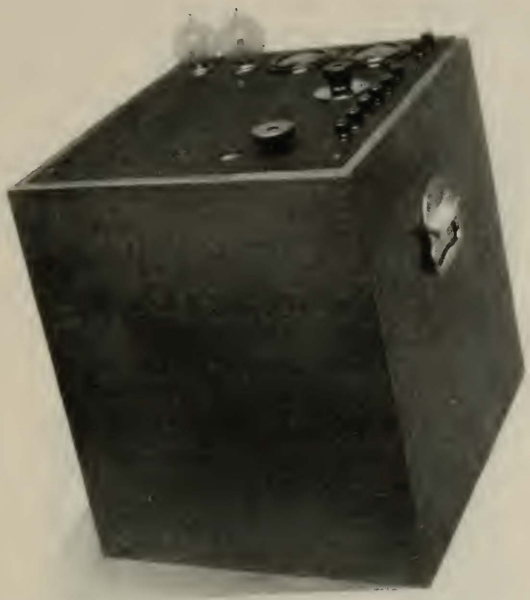


Fig. 2—Single-frequency vacuum tube oscillator used as secondary standard of frequency

The meter is calibrated by balancing the circuit by means of the variable resistance and capacitance with a known frequency input, and recording the coil and condenser settings. It is used for checking frequencies by reversing the process, that is, connecting the source of unknown frequency to the bridge, balancing as before, and determining the frequency by reference to the calibration. There are no input or output transformers connected to this circuit and on this account certain precautions must be taken in connecting the output and input circuits to it; but it is a relatively low impedance circuit, and troubles due to this cause have not been found serious.

Resistance. A convenient secondary standard of resistance is a dial box having the resistance units designed to meet the same requirements as the prime standards. Commercial dial boxes are available, having satisfactory stability with variations in frequency and atmospheric conditions, and having sufficiently small phase angles for all frequencies but the highest radio frequencies.

A dial box, requiring, as it does, a certain amount of wiring between dials, and having all of the dials connected permanently whether they

are used or not, always has more capacitance and inductance associated with it than a single resistance of the same value. A certain amount of compensation between the capacitance and inductance may be effected by proper design, but it may be generally accepted that the inductance of the wiring makes the phase angle of the low



Fig. 3—Resonance-type frequency meter

resistance values comparatively high and the capacitance between dials and between units of each dial makes the phase angle of the high values comparatively high. This effect can only be overcome by a compact design using coils of small physical size. This sets a limitation on coils for use in dial boxes which is not present to such an extent in the case of single resistance units or single value prime standards.

Methods of Measurement

We have discussed already measurements of frequency and resistance in connection with the description of standards, and we will not discuss them further here. We are particularly concerned with the measurement of impedance of all types, it being understood that any resistance having a phase angle which is not negligible or which is of special interest is to be considered a special type of impedance.

In measuring impedances, we have found that those methods which determine the unknown in terms of circuit constants are superior to those requiring the measurement of current and voltage. Accordingly, bridge methods are used almost exclusively and, furthermore, the bridge type which is used wherever possible is the equal ratio arm bridge in which a direct comparison is made of the unknown impedance with a known impedance adjusted to that same value. This type of measurement has the disadvantage of requiring standards of the same value as the quantity measured over the whole range of impedances used, but it has the compensating advantages that, having standards whose value is known, this circuit is extremely simple, very easy to check at any time, and may be made extremely accurate.

Auxiliary Apparatus. Without going into details regarding the auxiliary apparatus used in connection with bridge measurements, we may state briefly that vacuum tube oscillators are used almost exclusively for furnishing all frequencies, and that the telephone receiver is used almost exclusively as a detector, due to its simplicity and the rapidity with which it may be used. For frequencies below 200 cycles, it is used with a chopper to give a tone of about 1000 cycles, and above 3000 cycles, it is used with a heterodyne detector to give a beat note of about 1000 cycles. In the audio frequency range, it is used alone or with an amplifier, if necessary.

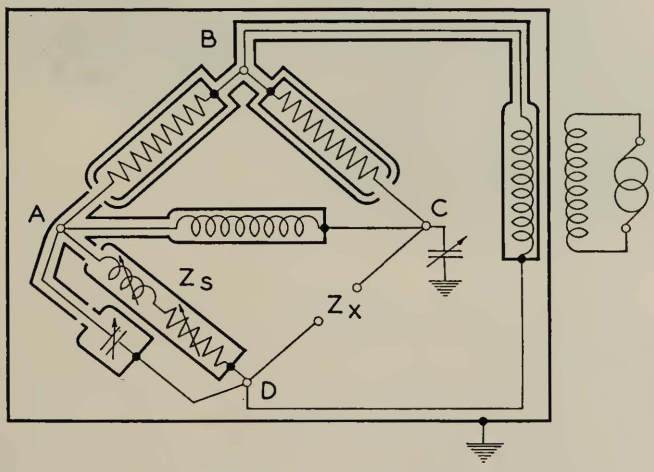


Fig. 4—Shielded impedance bridge circuit

While it is impossible to draw a distinct line between the methods of measurement of different types of impedances, certain bridge circuits have been designed primarily for certain types of measure-

ments, and we will therefore classify them in this way, although in general they have a considerably wider sphere of usefulness than indicated.

Inductance. A simple shielded bridge for the measurement of inductance and resistance has been described by one of the authors [4] and is shown in schematic form in Fig. 4. It comprises two equal resistance ratio arms, an adjustable standard of self-inductance, an adjustable resistance standard, a thermocouple milliammeter, two reversing switches, two transformers, and two air condensers. This apparatus is grouped into three separate units, as shown in Fig. 5, one comprising the

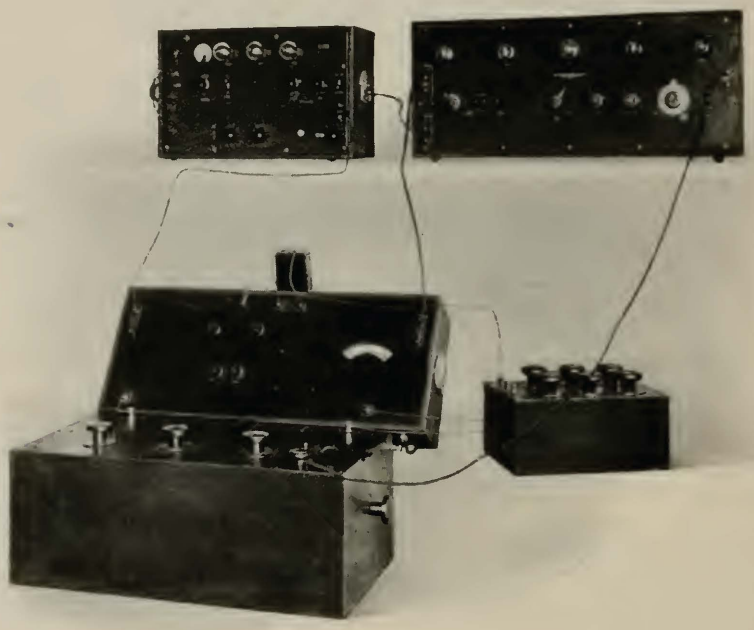


Fig. 5—Shielded impedance bridge and standards, connected to vacuum tube oscillator and heterodyne detector

standards of inductance, one the resistance standard, and one the remaining parts of the circuit. Each of these units is shielded electrostatically. The last assembly constitutes the balance element of the system, by means of which the unknown and standard impedances are compared. This unit may be used alone for the comparison of two impedances of any type since the only condition for balance is the exact equality of impedances in the two arms. Using in addition the standard inductance and resistance shown, it is adapted particularly for measuring inductance and effective resistance. The inductance

standard may be made with a range of 10 henrys to a minimum of two millihenrys, using an inductometer having a minimum scale division of 0.1 millihenry, or the range may be any simple multiple of this. Values as low as one microhenry at frequencies as high as 150 kc. are measured in this way.

By connecting the resistance in one arm of the bridge and a capacitance in series with an inductance in the other arm, we may use it to indicate resonance, and if we measure the frequency we may use this method for the comparison of capacitance with inductance. This is the method actually used for the calibration of the inductance standard used with the bridge. The bridge may be used for the comparison of capacitance. The bridge described later for the measurement of capacitance, however, has certain special features which make it peculiarly adapted to the measurement of capacitance and conductance.

Inductance with Superposed Direct Current. In telephone work, it is often of value to know the performance of apparatus, particularly of iron core impedances, when used at telephone frequencies while at the same time carrying direct current. The bridge shown in schematic form in Fig. 6 will measure the inductance of the coil at audio frequency

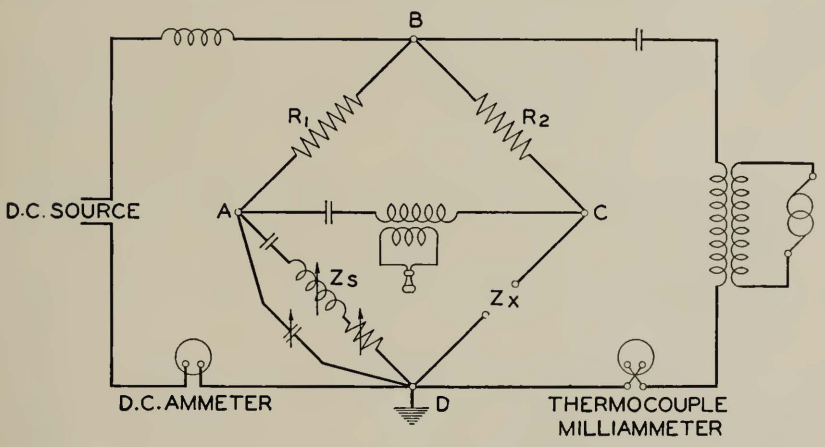


Fig. 6—Bridge circuit for measuring impedances with superposed direct current

with a direct current flowing through it. As shown in the figure, the direct current is kept out of all of the arms of the bridge except one ratio arm and the test arm, by means of condensers, and the alternating measuring current is separated from the direct current by means of a choke coil. None of these added features affect the bridge balance except the capacitance in the standard arm, and this is made large

enough (26 μ f.) to have an impedance small compared with the impedance measured. In any case, a correction may be made by taking first a zero reading which will be slightly positive due to the inductance necessary to compensate for the capacitance in this circuit. This correction will vary with frequency but at 1800 cycles, for instance, with 26-μ f. capacitance, the correction is only about 0.3 millihenry and the inductances measured are usually considerably larger than this.

The circuit is extremely simple and convenient to use. The values of alternating current and direct current can each be measured separately outside of the bridge circuit and the inductance standards do not need to be constructed to carry the direct current. The only part of the bridge required to carry the direct current is one ratio arm and, in consequence, it is a comparatively simple matter to construct such a bridge to carry several amperes of direct current. Where very high direct currents are required, the ratio arms may be reactances wound on a single core, instead of resistances, thus reducing the loss due to the passage of the direct current.

Flutter. In telephone circuits used for joint telephone and telegraph service, it is desirable to know the effect of the telegraph impulse on the telephone frequency inductance and effective resistance of the loading coils used on the lines. This effect, known as "flutter," with a method of measuring it, is described in detail by Fondiller and Martin.[7] The measuring circuit consists of a double bridge, the inner one consisting of two similar loading coils on which the flutter effect is to be measured and two other coils of comparatively high impedance approximately equal in value and which have negligible flutter effects, the four coils being connected to form a balanced bridge. The low frequency corresponding to the telegraph impulse is introduced at two diagonal corners and the other two corners, which are at a common potential with respect to the low frequency, are connected to the usual test terminals of an impedance bridge of the type already described. With no low-frequency current passing through the coils, a continuous balance may be obtained on the main or high-frequency bridge using an audio frequency input. From this, the normal effective resistance and inductance of the coils may be obtained.

When the low-frequency current passes through the coils, the inductance and effective resistance are different for every point of the low-frequency cycle. Thus, only an instantaneous balance of the outer bridge is possible. This instantaneous balance for any particular point in the low-frequency cycle may be made by the use of an electro-

[7] W. Fondiller and W. H. Martin, *Transactions of the A. I. E. E.*, 1921, Vol. 40, p. 553.

magnetic oscillograph. By this means as described in the paper already mentioned, it is possible to obtain the curve of variation of inductance and effective resistance of the coil over one low-frequency cycle.

Another method used at the present time employs the same bridge circuit but an entirely different method of detecting the cyclic variation in the balance. This method of detection uses the cathode-ray oscillograph and is as follows. The low-frequency source is connected across a high resistance and condenser in series, the two having equal impedances. The potentials across the condenser and resistance are then placed respectively across the horizontal and vertical plates of the oscillograph. These two potentials, being equal in magnitude but 90 deg. apart in phase, give a circle on the screen. The output of the main bridge is now connected through a transformer whose secondary is connected in series with the oscillograph cathode potential. Due to the fact that the sensitivity of the tube to deflections by the plate potentials varies with the cathode potential, the radius of this circle produced by the low frequency is a function of the telephone frequency input from the bridge, and instead of a circle we get a band, the width of which is a measure of the degree of unbalance of the bridge. The point in the cycle at which the bridge is balanced, is indicated on the screen as the point where this band diminishes to a line, and the angular position of this point in the band determines the phase position of this balance with respect to the low-frequency cycle. It is possible in this way to balance the bridge for any angular position corresponding to any point in the low-frequency cycle, and by taking sufficient points, to obtain a curve of variation of the coil constants over a complete cycle. This method is found to be simpler and faster than the method using the mechanical oscillograph.

Inductance Balance. A simple form of bridge for measuring inductance balance of the two windings of a transformer or other coil uses the two windings of the transformer for two arms, the other two arms being resistances, one of which at least is variable. The balance is made by means of the variable resistance, the ratio of the two resistances at balance then giving the unbalance of the transformer. If one of these resistances is made 100 ohms, the variation of the other from 100 ohms at balance gives directly the percentage unbalance. Any unbalance in resistance is usually comparatively small and may be taken care of by low resistances in series with the transformer windings.

Ratio of Transformation. A similar bridge may be used for the measurement of ratio of transformation. There are many cases where

the secondary of a step-up transformer has an inductance which is inconveniently large to measure directly, and the ratio of transformation circuit eliminates this necessity. The circuit used is practically the same as that already described for measuring inductance balance, the ratio of transformation being equal to the ratio of the resistance arms of the bridge at balance.

Capacitance. The direct comparison of capacitance is made in a special bridge known as the Campbell [8]-Colpitts [9] capacitance and conductance bridge. The ratio arms, input and output circuits, and the shielding are similar to the impedance bridge already described. The unique feature of this bridge is the method of connecting the standard air condenser to eliminate the dielectric loss in the measurement of capacitance. The schematic diagram of the bridge is shown in Fig. 7. Instead of connecting the standard condensers in the

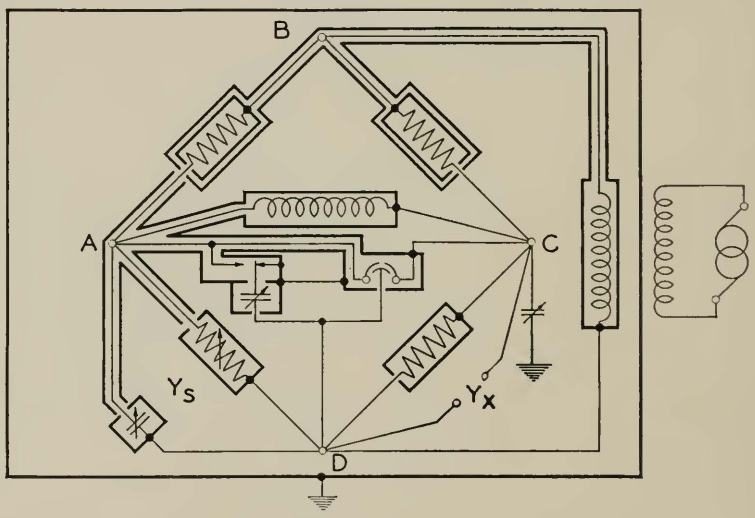


Fig. 7—Schematic circuit of capacitance and conductance bridge

arm *AD* as in the case of the impedance bridge already described, a special switch is used to switch these condensers from *AD* to *CD*, and in the case of the continuously variable condenser, the three-plate construction is used, causing a decrease in the capacitance in *CD* as the capacitance in *AD* is increased.

The method of construction of the unit air condensers is shown in

[8] G. A. Campbell, "The Shielded Balance," *Electrical World and Engineer*, April 2, 1904, p. 647.

[9] G. A. Campbell, "Measurement of Direct Capacities," *Bell System Technical Journal*, July, 1922, p. 18.

Fig. 8. It may be seen from this figure that all capacitances which include dielectric material are permanently connected across CD or AC and so are not changed when the condenser is switched, or else

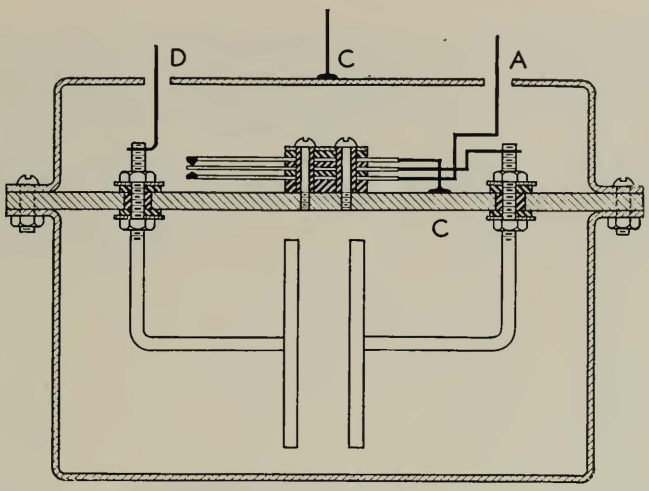


Fig. 8—Air-condenser construction employed in the capacitance and conductance bridge

they are switched so that capacitances across AC, which do not enter into the bridge balance, are short-circuited on switching. This scheme eliminates all dielectric loss in the standards when measuring condensers by comparison with them. It has the additional advantage that the capacitances in the bridge have twice the effect they would have if simply switched in and out of the circuit.

By the use of this bridge, it is possible to measure capacitances up to the maximum limit of the range of the air condensers with a negligible loss in the standard condensers. This capacitance range is usually up to 0.01 μ f. and for condensers above this value the conductance is measured by comparison with that of the maximum value of the air condenser, assuming it to have negligible conductance. Of course this method of eliminating dielectric loss is not applicable to the use of mica condenser standards and if a range greater than 0.01 μ f. is desired, the mica condensers are simply connected in the usual way across AD.

Another feature of this bridge is the method of measuring conductance. The connection of a variable resistance, either in series or in shunt, with the standard condenser for the measurement of loss in the test condenser has objections due to the wide range of resistance

values required to cover the possible variations in losses. A compromise is effected in this bridge by connecting a 10,000-ohm shunt across each of the arms *CD* and *AD*. A slight difference in the losses in these two arms can then be measured by varying one of these resistances slightly. Since the standard condenser practically always will have lower losses than the condenser tested, it is usual to place a fixed 10,000-ohm resistance across *CD* and a resistance across *AD* variable in 0.01-ohm steps to 10,000 ohms. A change of one ohm in this resistance, when balancing a condenser, is equivalent to shunting it with a resistance of 100 megohms or 0.01 micromho. Accordingly, the conductance of a condenser may be measured in micromhos by simply dividing the resistance change in ohms by 100. This, of course, is only approximate in the case of large conductances, but is correct to 1 per cent for values up to one micromho.

Due to the condensers forming such an integral part of the bridge circuit, they are all built into the bridge. The complete bridge is shown in Figs. 9 and 10. Fig. 9 is a top view showing the capacitance and



Fig. 9—Capacitance and conductance bridge

resistance dials for effecting a balance, and Fig. 10 is a view with the cover removed, showing the method of shielding the individual parts. The range of capacitance is from 0.1 $\mu\mu$ f. up to three μ f., and the frequency range is from about 10 cycles up to about 150 kc., the only modifications required in the bridges to cover this whole frequency

range being a change in input and output transformers, as it is not found practicable to design these transformers to give efficient operation over such a wide frequency range.

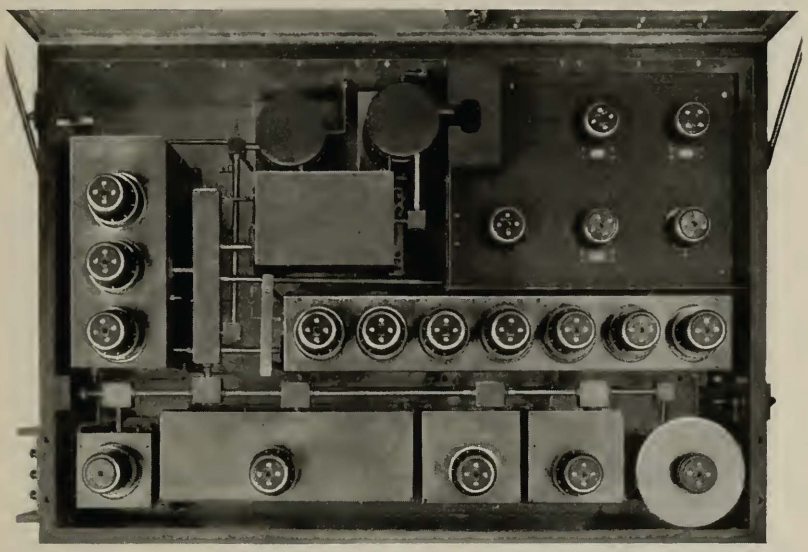


Fig. 10—Capacitance and conductance bridge with cover removed, showing method of assembly and shielding

A comparison of this bridge with the impedance bridge already mentioned shows it to be essentially the same circuit, the capacitance bridge having conductance shunts not included in the impedance bridge which allow a conductance balance to be made more readily. It is obvious that any two impedances can be compared on this bridge. Inductances may be measured by parallel resonance by simply placing them in the AD arm in parallel with the standard condenser and effecting a balance with it. This method is used to some extent for the measurement of large inductances.

Capacitance Unbalance. In order to keep cross-talk low in long cable circuits, it is necessary to have a high degree of capacitance balance between the various conductors in the cable, more particularly between the four conductors of a phantom group. The unbalances of interest are the phantom to each side circuit and the side-to-side unbalances. These may be measured on a capacitance bridge by measuring all of the direct capacitances [9] associated with the group and computing the unbalances required. A special circuit, however, is generally used which measures directly the particular un-

balances in which we are interested. It consists of an input and an output transformer, two equal resistance ratio arms, a variable air condenser of the three-plate type, four binding posts for connecting the four conductors of the quad, and switches for making the various connections. By means of the switches, the cable conductors are connected to the circuit in such a way that the reading of the air condenser when a balance is obtained indicates directly the unbalance, either side-to-side or phantom-to-side, according to the switch positions. This circuit when used as a laboratory instrument is capable of measuring capacitance unbalance as low as 1 $\mu\mu$ f.

Attenuation and Gain. So far, we have discussed the measurement of the fundamental impedance characteristics of apparatus. When the component parts have been found to meet their individual impedance requirements and are assembled to form the completed apparatus, it is desirable to have tests made of the over-all performance of this apparatus. In a large number of cases, the requirement of greatest importance is the attenuation frequency characteristic. It is fairly obvious that this characteristic, of all apparatus used in telephone lines, is of interest, and this is particularly true of all types of filter circuits which are designed primarily for the purpose of furnishing definite attenuation frequency characteristics. These measurements are particularly required on apparatus used in carrier-current telephony and telegraphy.

From the very nature of the measurements, it is difficult to obtain a null method of measuring attenuation. The most direct method is to measure the input and the output of the apparatus under test simultaneously, from which the attenuation may be computed. The practical difficulty in doing this is to measure the extremely small outputs which are obtained from apparatus having high attenuations, where the characteristic must be obtained with the normal input, which is usually low. In general, it has been found necessary to use some form of amplifying device in the output circuit and it has not been found desirable to rely on the constancy of amplification of this device. Accordingly, the usual method used for the measurement of attenuation is a substitution one. The circuit is shown in Fig. 11A. There are two branches in this circuit, one of which includes the apparatus under test and the other, a variable standard attenuator. The output of each branch is arranged to connect either to a detector of impedance Z_1 equal to the impedance of the standard attenuator or to a fixed impedance of the same value. If the apparatus under test has the same impedance as the standard attenuator, the input impedances Z_1 and Z_3 are made equal and the matching impedance Z_2

is omitted. Then the two branches of the circuit will be identical, provided the attenuation of the standard attenuator is equal to that of the apparatus under test. Accordingly, the method of measurement is to switch the detector first to one and then to the other branch,

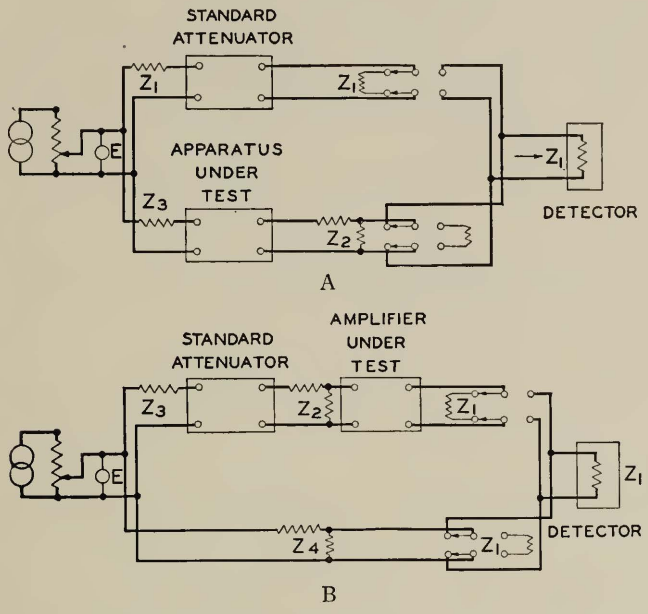


Fig. 11—Circuits for measuring attenuation and gain. A. Arrangement for measuring loss. B. Arrangement for measuring gain

adjusting the standard attenuator until an equal output is obtained for either switch position. The attenuator then reads directly the loss in the apparatus. The total input of the circuit is independent of the switch position, since the impedance conditions remain unchanged in switching.

If the apparatus under test has not the same impedance as the standard attenuator, the input impedance Z_3 and the matching network Z_2 are adjusted so that the circuit still reads directly.

The standard attenuator is a resistance network capable of variation in small steps, each step consisting of a network of the L, T or H type, the resistance values being such as to give the desired attenuation between the output and input terminals. It is usually calibrated in 0.1-T.U. steps and may read as high as 100 T.U. corresponding to a ratio of power output to power input of ten billion to one or, if the impedances are the same, which is usually the case, corresponding to a current or voltage ratio of 100,000 to 1.

The calibration of these attenuators is based on the measurement of the individual resistances. Of course, sufficient measurements are made to determine that any capacitances which enter do not affect appreciably the accuracy of the attenuator at the maximum frequency used, which may be as high as 150 kc.

By modifying the circuit of Fig. 11A, we may use it to measure gain as shown in Fig. 11B. In this arrangement, the lower branch contains an impedance Z_4 that is adjusted to introduce a loss equal to that of the matching impedance Z_2 in the upper branch. In other words, with the amplifier under test out of the circuit and the standard attenuator set at zero, the detector will read the same for either position of the output switch. Then when the amplifier is introduced into the circuit, the attenuator is adjusted until the detector reads the same for either switch position, which means that the gain of the amplifier is just neutralized by the attenuator and the setting of the latter is read as gain.

This circuit is used principally for the measurement of gain of audio frequency amplifiers, and is capable of measuring gain as high as 120 T.U. corresponding to a power output of 1,000,000,000,000 times the power input.

Cross-Talk. When there is an appreciable amount of coupling between two telephone circuits, any mutual interference which results is known as cross-talk. It is measured in cross-talk units, a cross-talk unit being defined as the relation existing between the two circuits when the current in the disturbed circuit is one millionth of the current in the disturbing circuit, the impedances of the two circuits being the same. Under these conditions, one cross-talk unit may be assumed the same as 120 T.U. An interesting form of cross-talk is that due to loading coils and is of a complex type, produced by a combination of capacitance, inductance and resistance unbalances in the windings. Since the actual cross-talk caused by an unbalance in the coil is dependent upon all of the conditions of the circuit, it is necessary that any measurement of cross-talk made on the individual coils be made in a circuit as nearly as possible the equivalent of the line in which the coil is to be used. Consequently, all cross-talk circuits for the measurement of loading coil cross-talk consist of networks simulating the impedance of an ideal line of the type for which the loading coil is designed. The principle of the method is to apply to the disturbing circuit a definite input of a single frequency, usually 900 cycles, and to measure the cross-talk in the disturbed circuit at the desired point in it by comparing the tone heard in the telephone receiver connected at this point with the tone obtained from a cross-talk meter which is simply a device

for obtaining a definite part of the input, and having a scale reading in millionths, that is, in cross-talk units. The measurement is made by switching from the cross-talk meter to the disturbed line and adjusting the cross-talk meter until the tone heard in each case is the same. The method is therefore not a null method and depends to some extent on the judgment of the operator, but results accurate to one or two cross-talk units may be obtained by this method. The coils as commercially produced after adjustment for this requirement are usually within 10 cross-talk units, representing an unbalance in the circuit due to the coil unbalance of less than one part in 100,000.

Conclusion

We have described in this paper a number of the more important high-frequency methods of measurement and measuring circuits. It has been impossible to cover all of the different methods and circuits used, but we believe that the information given will be of value to those interested in this field of work.

We have not been able, in a paper of this type, to go into details concerning any specific circuits used, but we have referred to papers which describe in greater detail some of these methods and circuits, and it is expected that other papers will be published in the future covering other circuits which have received only brief mention here.

The Diffraction of Electrons by a Crystal of Nickel

By C. J. DAVISSON

This article is taken from the manuscript prepared by the author for his address at the joint meeting of Section B of the American Association for the Advancement of Science and the American Physical Society on December 28, 1927, at Nashville, Tennessee. An account of this work giving fuller experimental details is given by Davisson and Germer in the December, 1927, issue of the *Physical Review*.

These experiments are fundamental to some of the newer theories in physics. Until they were performed, it could be said that all experimental facts about the electron could be explained by regarding it as a particle of negative electricity. It now appears that in some way a "wave-length" is connected with the electron's behavior. The work thus shows an interesting contrast with the discovery of A. H. Compton that a ray of light (a light pulse) suffers a change of wave-length upon impact with an electron, the change of wave-length corresponding exactly to the momentum gained by the electron. Until Compton's work, all the known facts about light could be explained by thinking of light as a wave motion. The Compton effect seems to prove the existence of particles of light.

Physics is thus faced with a double duality. Compton showed that light is in some sense *both* a wave motion and a stream of particles. Davisson and Germer have now shown that a beam of electrons is in some sense *both* a stream of particles and a wave motion.

At the same time, theoretical advances have been made which seem to pave the way for an understanding of this curious situation. A general account of these new developments was given by K. K. Darrow in his series "Contemporary Advances in Physics" in the *Bell System Technical Journal* for October, 1927. Some remarks on the relation of the Davisson and Germer experiments to the new mechanics were given in this article, p. 692 *et seq.*—EDITOR.

THE experiments which I have been asked to describe are the most recent of an investigation of the scattering of electrons by metals on which we have been engaged in the Bell Telephone Laboratories for the last seven or eight years.

The investigation had its inception in a simple but significant observation. We observed some time in the year 1919 that when a beam of electrons is directed against a metal target, electrons having the same speed as those in the incident beam stream out in all directions from the bombarded area. It seemed to us at the time that these could be no other than particular electrons from the incident beam that had suffered large deflections in simple elastic encounters with single atoms of the target. The mechanism of scattering, as we pictured it, was similar to that of alpha ray scattering. There was a certain probability that an incident electron would be caught in the field of an atom, turned through a large angle, and sent on its way without loss of energy. If this were the nature of electron scattering it would be possible, we thought, to deduce from a statistical study of the deflections some information in regard to the field of the

deflecting atom. It was with these ideas in mind that the investigation was begun. What we were attempting, it will be seen, were atomic explorations similar to those of Sir Ernest Rutherford and his collaborators but explorations in which the probe should be an electron instead of an alpha particle. I shall not stop to recount the earlier experiments of this investigation, but shall pass at once to the most recent ones—those in which Dr. Germer and I have studied the scattering of electrons by a single crystal of nickel.

The unusual interest that attaches to these experiments is due to their revealing the phenomenon of electron scattering in a new and, I may say, fashionable rôle. Electron scattering is not, it would seem, the mildly interesting matter of flying particles and central fields that we supposed, but is instead a much more interesting phenomenon in which electrons exhibit the properties of waves. The experiments reveal that the way in which electrons are scattered by a crystal is very similar to the way in which x-rays are scattered by a crystal. The analogy is not so much with the alpha ray experiments of Sir Ernest Rutherford, as with the x-ray diffraction experiments of Professor von Laue.

My task of describing these experiments is much simplified by the fact that the experiments of Professor von Laue are so well known and so thoroughly comprehended. I remind you very briefly that in the original Laue experiment a beam of x-rays was directed against a crystal of zincblende, that about the transmitted beam was found an array of regularly disposed subsidiary beams proceeding outward from the irradiated portion of the crystal, and that these subsidiary beams could be interpreted completely and precisely in terms of the then already popular wave theory of x-radiation. They could indeed be explained as diffraction beams that resulted from the superposition of secondary wave trains expanding from the regularly arranged atoms of the crystal lattice.

There are two features of the Laue experiment which we shall need particularly to remember. The first is that diffraction beams issue not only from the far side of the crystal along with the transmitted beam, but also from the near or incidence side of the crystal—these latter being disposed in a regular array about the incident beam. The second is that each diffraction beam is characterized by a particular wave-length, and that a given beam appears in the diffraction pattern if the incident beam contains radiation of its characteristic wave-length, or of some submultiple value of this wave-length, but not otherwise. If the incident beam is monochromatic, no diffraction beams appear at all unless the wave-length of the incident beam

happens to coincide with a wave-length of one or more of the diffraction beams. In that case the favored beams appear but no others.

With this picture of x-ray scattering in mind one sees at once the significance of the main results of the present experiments. A homogeneous beam of electrons is directed against a crystal of nickel, and at certain critical speeds of bombardment full speed scattered electrons issue from the incidence side of the crystal in sharply defined beams —a few beams at each of the critical speeds—the totality of such beams making up a regularly disposed array similar to the array of Laue beams that would issue from the same side of the same crystal if the incident beam were a beam of x-rays.

The electron beams are not identical in disposition with the Laue beams, and yet it is possible to treat them as diffraction beams, and from their position and from the geometry and scale of the crystal to calculate "wave-lengths" of the incident beam—just as we might do if we were dealing with x-rays or with any other wave radiation. When this is done we arrive at a definite and simple relation between the speed of the electron beam and its apparent wave-length—the wave-length is inversely proportional to the speed.

Surprising as it is to find a beam of electrons exhibiting thus the properties of a beam of waves, the phenomenon is less surprising today than it would have been a few years ago. We have been prepared, to a certain extent, by recent developments in the theory of mechanics for surprises of just this sort—for the discovery of circumstances in which particles exhibit the properties of waves. We have witnessed, during the last three years, the inception and development of the idea that all mechanical phenomena are in some sense wave phenomena—that the rigorous solution of every problem in mechanics must concern itself with the propagation and interference of waves. The wave nature of mechanical phenomena is not ordinarily apparent, we are told, because the length of the waves involved is ordinarily small compared to the dimensions of the system. It is only in such small scale phenomena as the intimate reactions between atoms and electrons that the wave-lengths are comparable with the dimensions of the system. Here only are we to expect notable departures from classical mechanics, and here only are we to find evidence of a more comprehensive wave mechanics.*

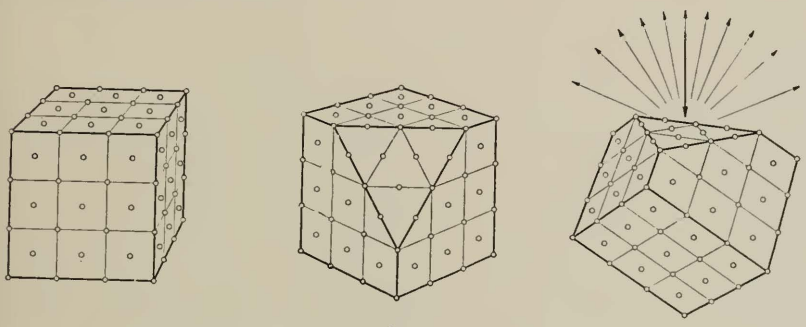
The success of this new theory has been confined, up to the present time, to explanations of certain of the data of spectroscopy. In this field the theory has appealed very strongly to all of us because of the

* It was predicted by W. Elsasser in 1925 (Naturwiss., *13*, 711 (1925)) that evidence for the wave mechanics would be found in the interaction between a beam of electrons and a crystal.

elegance of its methods and because of its remarkable facility in accounting for various of the inhibitions with which the radiating atom is afflicted. We have been prepared by these successes to view with not too great surprise—or alarm—evidence for the wave nature of phenomena involving freely moving electrons. And any reluctance we may feel in treating electron scattering as a wave phenomenon is apt to be dispelled when we find that the value calculated for the wave-length of the equivalent radiation is in acceptable agreement with that which L. de Broglie assigned to the waves which he associated with a freely moving particle—that is to say, the value h/mv (Planck's constant divided by the momentum of the particle).

In this account of the experiments I will describe the general method of the measurements and the general character of the results rather than attempt to go into these matters in detail.

Nickel forms crystals of the face centered cubic type. In Fig. 1 (*a*) the crystal which we had at our disposal is represented by a block of unit cubes of this type.



a *b* *c*

Fig. 1—Diagrams of nickel lattice, of cut lattice, and of lattice with incident and scattered beams

Our first step in preparing the crystal for bombardment was to cut through this structure at right angles to one of the cube diagonals. The appearance of the crystal after the cut was made, and the corner of the cube removed, is indicated in Fig. 1 (*b*). It is this newly formed triangular surface that was exposed to electron bombardment. The bombardment was at normal incidence as indicated in Fig. 1 (*c*). We are to think of electrons raining down normally upon this triangular surface, and of some of these emerging from the crystal without loss of energy, and proceeding from it in various directions.

What is measured is the current density of these full speed scattered electrons as a function of direction and of bombarding potential. The way in which the measurements are made is illustrated in Fig. 2. The electrons proceeding in a given direction from the crystal

enter the inner box of a double Faraday collector and a galvanometer of high sensitivity is used to measure the current to which they give rise. An appropriate retarding potential between the parts of the collector excludes from the inner box all but full speed electrons.

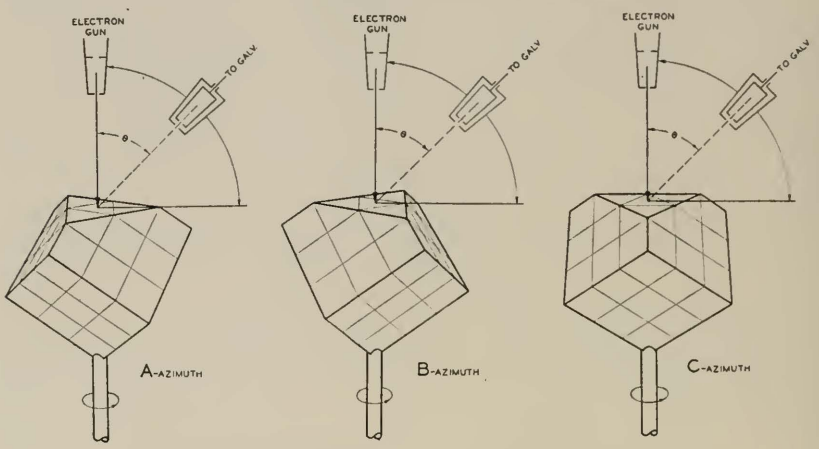


Fig. 2—Showing the three principal azimuths

The collector may be moved over an arc of a circle in the plane of the drawing as indicated, and the crystal may be rotated about an axis which coincides with the axis of the incident beam of electrons. Thus the collector may be set for measuring the intensity of scattering in any direction relative to the crystal—by turning the crystal to the desired azimuth, and moving the collector to the desired colatitude. The whole solid angle in front of the crystal may be thus explored with the exception of the region within twenty degrees of the incident beam.

Certain of the azimuths related most simply to the crystal structure we shall refer to as "principal azimuths." Thus there are the three azimuths that include the apexes of the triangle. If we find the intensity of scattering depending on colatitude in a certain way in one of these azimuths, we expect, of course, to find it depending upon colatitude in the same way in each of the other two. We shall call these the "A-azimuths." On the left in Fig. 2 the crystal has been turned to bring one of the A-azimuths into the plane of rotation of the collector.

Another triad of principal azimuths consists of the three which include the mid-points of the sides of the triangle. These we shall call the "B-azimuths." The next most important family of azimuths comprises those which are parallel to the sides of the triangle; of these there are six, the "C-azimuths."

If we turn the crystal to any arbitrarily chosen azimuth, set the bombarding potential at any arbitrarily chosen value, and measure the intensity of scattering as a function of colatitude, what we find ordinarily is the type of relation represented by the curve on the left in Fig. 3.

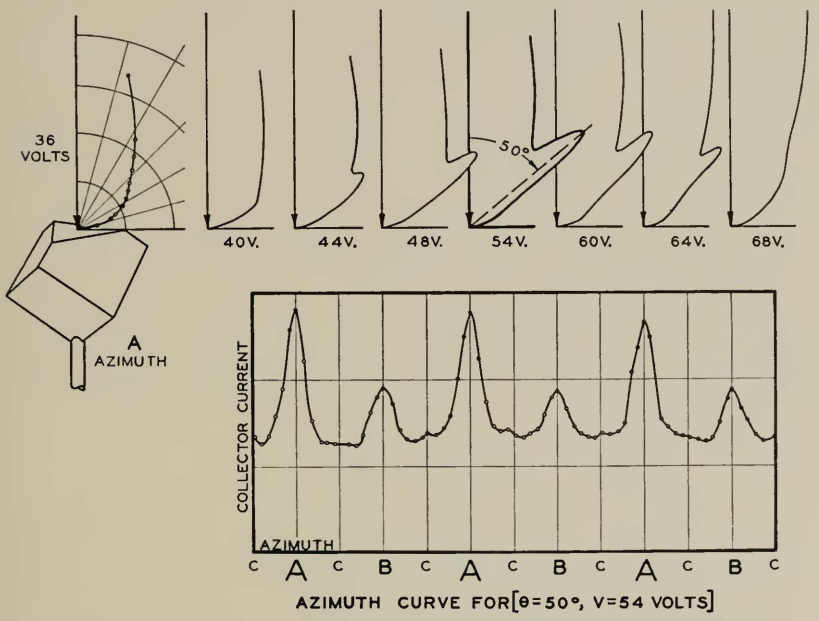


Fig. 3—Curves showing development of diffraction beam in the A-azimuth . . . and variation of intensity with the azimuth at colat. 50° for which beam is strongest in the A-azimuth

This curve is actually one found for scattering in the A-azimuth when the bombarding potential is 36 volts. It is typical, however, of the curves that are obtained when no diffraction beam is showing. The intensity of scattering in a given direction is indicated by the length of the vector from the point of bombardment to the curve. The intensity is zero in the plane of the crystal surface, and increases regularly as the colatitude angle is decreased. This type of scattering forms a background upon which the diffraction beams are superposed.

The occurrence of a diffraction beam is illustrated in the series of curves to the right in Fig. 3. When the bombarding potential is increased from 36 to 40 volts, the curve is characterized by a slight hump at colatitude 60 degrees. With further increase in bombarding potential this hump moves upward, and at the same time develops

into a strong spur. The spur reaches its maximum development at 54 volts in colatitude 50 degrees, then decreases in intensity, and finally vanishes from the curve at about 70 volts in colatitude 40 degrees.

We next make an exploration in azimuth through this spur at its maximum; we adjust the bombarding potential to 54 volts, set the collector in colatitude 50 degrees, and make measurements of the intensity of scattering as the crystal is rotated. The results of this exploration are exhibited by the curve at the bottom of Fig. 3, in which current to the collector is plotted against azimuth. We find that the spur is sharp in azimuth as well as in latitude and that it is one of a set of three spurs as required by the symmetry of the crystal.

We observe also that there are small spurs showing in the B-azimuths. We turn the crystal to bring the B-azimuth under observation, and again make explorations in latitude for various speeds of bombardment. We find that the spur in the B-azimuth is similar to the "54 volt" spur in the A-azimuth, but that it attains its maximum development at a higher voltage and at a higher angle. Curves exhibiting its growth and decay are shown in Fig. 4. Maximum

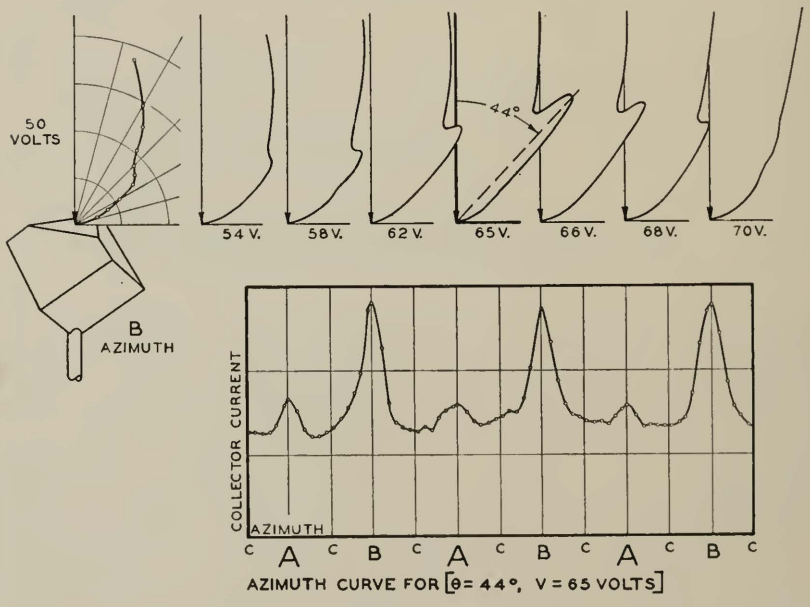


Fig. 4—Similar for the B-azimuth

development is attained at 65 volts in colatitude 44 degrees. At the bottom of the figure we show the intensity-azimuth curve through

this spur at its maximum. The small maxima in the A-azimuths represent the remnants of the "54-volt" spurs.

We have thus a set of spurs at colatitude 50 degrees in the A-azimuths when the bombarding potential is 54 volts and a set of 44 degrees in the B-azimuths when the bombarding potential is 65 volts. These spurs are due to beams of full speed scattered electrons which are comparable in sharpness and definition with the beam of incident electrons. This is inferred from the widths of the spurs and the resolving power of the apparatus.

It is hardly necessary to point out that these sharply defined beams of scattered electrons are similar in their behavior to x-ray diffraction beams. If the incident beam were a beam of monochromatic x-rays of adjustable wave-length instead of a homogeneous beam of electrons of adjustable speed, quite similar effects could be produced. If the wave-length of the x-ray beam were varied, critical values would be found at which intense diffraction beams would issue from the crystal in its A-azimuths and others at which such beams would issue in the B-azimuths. The x-ray diffraction beams would indeed be more sharply defined in wave-length than the electron beams defined in voltage. No diffraction beam would be observed until the wave-length of the incident x-rays were very close indeed to its critical value, and the beam would disappear again when the wave-length had passed only very slightly beyond the critical value. This "wave-length sharpness" or "wave-length resolving power" is dependent, however, upon the number and disposition of the atoms involved in the diffraction. If the crystal were only a few atom layers in thickness, or if the x-rays were extinguished on penetrating through only a few atom layers of the crystal, then the x-ray diffraction beams would be much less sharply defined in wave-length; they would behave more like the electron beams. We may say then that the electron beams exhibit the general behavior of diffraction beams resulting from the scattering of a beam of very soft wave radiation—radiation that is very rapidly extinguished in the crystal.

Let us try now to forget that what we are measuring in these experiments is a current of discrete electrons arriving one by one at our collector. Let us imagine that what we are dealing with is indeed a monochromatic wave radiation, and that our Faraday box and galvanometer are instruments suitable for measuring the intensity of this radiation. We are to think of the incident electron beam as a beam of monochromatic waves, and of the "54-volt beam" in the A-azimuth and the "65-volt beam" in the B-azimuth as diffraction beams that owe their intensities, in the usual way, to constructive

interference among elements of the incident beam scattered by the atoms of the crystal. With this picture in mind we try next to calculate wave-lengths of this electron radiation from the data of these beams and from the geometry and scale of the crystal.

To begin with, we shall need to look more closely into our crystal. The atoms in the triangular face of the crystal may be regarded as arranged in lines or files at right angles to the plane of the A- and B-azimuths. If a beam of radiation were scattered by this single layer of atoms, these lines of atoms would function as the lines of an ordinary line grating. In particular, if the beam met the plane of atoms at normal incidence, diffraction beams would appear in the A- and B-azimuths, and the wave-lengths and inclinations of these beams would be related to one another and to the grating constant d by the well-known formula, $n\lambda = d \sin \theta$, as illustrated at the top of the figure.

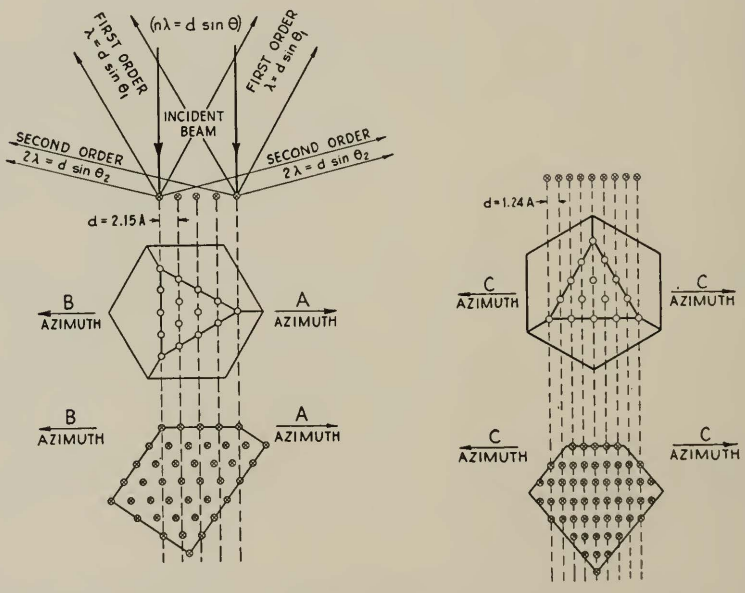


Fig. 5—Showing $n\lambda = d \sin \theta$ relation in the A-, B- and C-azimuths

In the actual experiments the diffracting system is not quite so simple. It comprises not a single layer of atoms, but many layers; it is equivalent not to a single line grating, but to many line gratings piled one above the other, as shown graphically at the bottom of the figure. What diffraction beams will issue from this pile of similar and similarly oriented plane gratings?

The answer to this question is twofold. In respect of position all the beams which appear will coincide with beams which would issue from a single grating. We get no additional beams by adding extra layers to the lattice. In respect of intensity, however, the results are greatly changed. A given beam may be accentuated or it may be diminished, both absolutely and relatively to the other beams; it may in fact be blotted out completely, or reduced to such an extent that it can no longer be perceived. These are effects of interference among the similar beams proceeding from the various plane gratings that make up the pile. Later we shall consider under what conditions these component beams combine to produce a resultant beam of maximum intensity; for the present, however, I wish only to stress the fact that whenever and wherever a space lattice beam appears its wave-length and colatitude angle θ will be related to the constant d of the plane grating through the ordinary plane grating formula. We therefore apply this formula to the 54- and 65-volt beams that have been described. The grating constant d has the value 2.15 Å., the 54-volt beam occurs at $\theta = 50°$ so that $n\lambda$ for this beam should have the value $2.15 \times \sin 50°$, or 1.65 Å. For the 65-volt beam we obtain for $n\lambda$ the value 1.50 Å.

We now compare these wave-lengths with the wave-lengths associated with freely moving electrons of these speeds in the theory of wave mechanics. Translated into bombarding potentials, de Broglie's relation

$$\lambda = h/mv \text{ becomes } \lambda = \sqrt{\frac{150}{V}} \text{ Å.},$$

where V represents the bombarding potential in volts. The length of the phase wave of a "54-volt electron" is $(150/54)^{1/2} = 1.67$ Å., and for a 65-volt electron 1.52 Å. The 54- and 65-volt electron beams do very well indeed as first order phase wave diffraction beams.

It may be mentioned that beams occur at different voltages in the A- and B-azimuths because the plane gratings that make up the crystal are not piled one immediately above the other. There is a lateral shift from one grating to the next amounting to one third of the grating constant. Because of this shift the phase relation among the elementary beams emerging in the A-azimuth is not the same as that among those emerging in the B-azimuth—and coincidence of phase among these beams occurs at different voltages, or at different wave-lengths, in the two azimuths.

We next make similar calculations for a beam occurring in the C-azimuth. One such beam attains its maximum development in

colatitude 56° when the bombarding potential is 143 volts. For diffraction into the C-azimuth we must regard the atoms in the surface layer as arranged in lines normal to the plane of this azimuth as illustrated in Fig. 5. The grating constant is 1.24 Å., and the similar gratings that make up the whole crystal are piled up without lateral shift. For this reason the C-azimuth is six-fold instead of only three-fold. For a beam occurring in this azimuth in colatitude 56°, $n\lambda$ should be equal to 1.24 × sin 56° or 1.03 Å. The value of h/mv for electrons that have been accelerated from rest through 143 volts is $(150/143)^{1/2}$ or 1.025 Å. Again the beam does very well as a first order diffraction beam.

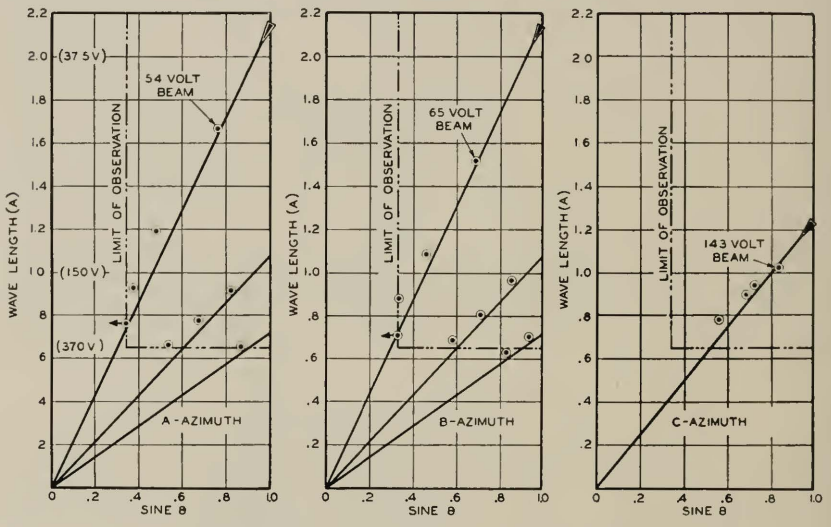


Fig. 6—Plot of λ against sin θ for various beams

The total number of such beams which we have observed in all azimuths in explorations up to 370 volts is twenty-four—nine in the A-azimuth, ten in the B-azimuth, and five in the C-azimuth. It would be possible to calculate an observed wave-length for each of these beams from $n\lambda = d \sin \theta$, and to compare this in each case with the theoretical wave-length calculated from $\lambda = h/mv$, just as we have done already for three of the beams. We have chosen, however, to display the results graphically rather than numerically.

The data for the twenty-four beams are exhibited in diagrams in Fig. 6, in which wave-length λ is plotted against the sine of the co-

latitude, θ. There is a separate diagram for each azimuth and in each the straight lines passing through the origin represent the plane grating formula $n\lambda = d \sin \theta$ in its different orders. Each of the twenty-four beams is represented by a point or by a wedge-shaped symbol in one of these diagrams. The quantities coordinated in each case are the wave-length of the incident beam as calculated from $\lambda = h/mv = (150/V)^{1/2}$ and the sine of the colatitude angle of the diffraction beam as observed. There are no points to the left of the line $\theta = 20°$ as this represents the lower limit of our colatitude range of observation, and none below the line $\lambda = 0.637$ Å. as this corresponds to the upper limit of our voltage range, 370 volts. The bombarding potentials corresponding to various wave-lengths are shown by figures enclosed in brackets.

When the data are exhibited in this fashion the question as to whether or not the observed wave-length of a beam agrees with its theoretical wave-length is answered by whether or not the point representing the beam falls on one or another of the lines representing the plane grating formula. If there were perfect agreement in all cases, each of the points would lie on some one of these lines.

It will be seen that the points all lie close to the lines, though not as a rule exactly on them. It is of course very important to decide whether the departures of the points from the lines are or are not too great to be attributed to uncertainties of measurements. It is our belief that they are in fact due to experimental error in the determination of the colatitude angles. If we accept the theoretical values of the wave-lengths as correct, and calculate the values of θ which we should have observed, we find that in no case do they deviate by more than 4 degrees from the values of θ actually set down. Corrections of this magnitude do not seem excessive when it is considered that we are making measurements with what amounts to a rather crude spectrometer, that the arm of the spectrometer is but 11 mm. in length, that the opening in the collector is 5 degrees in width, and that the spectrometer itself is sealed into a glass bulb. We therefore assume that in every case the value of the wave-length assigned by de Broglie is the correct one.

I now direct your attention to a particular group of these beams—the group comprising the beam of greatest wave-length in each of the three azimuths, which are represented in the figure by wedge-shaped symbols. The interpretation of these three is quite simple.

The radiation to which our electron beam is equivalent is extremely soft as already noted. Its intensity suffers a considerable decrement when the beam passes normally through only a single layer of atoms. This characteristic is inferred from the low resolving power of the crystal, and is consistent with what we know of the penetrating power of low speed electrons. When the beam passes through a layer of atoms at other than normal incidence the decrement in its intensity is greater still—and in the limit as the angle of incidence approaches grazing to the atom layer the intensity of the transmitted beam will approach zero. Thus we may expect that when a diffraction beam leaves the crystal at near grazing emergence the contributions to the resultant beam which come from the second and lower layers of atoms will be much less important than when the beam emerges from the crystal at a higher angle. Near grazing the radiation proceeding from the second and lower layers will be heavily absorbed in its passage through the overlying layers. Within a limited angular range near grazing the diffraction beam will be made up almost entirely of radiation scattered by the uppermost layer of atoms. The diffracting system becomes essentially a single plane grating and what we should observe is ordinary plane grating diffraction.

The first order diffraction beam from a line grating appears at grazing emergence when the wave-length of the incident radiation is equal to the grating constant. The grating constant for diffraction into the A- and B-azimuths is 2.15 Å. and grazing beams should appear in both azimuths when the wave-length of the incident electron beam has this value. The bombarding potential corresponding to wave-length 2.15 Å. is 32.5 volts, and at just 32.5 volts diffraction beams appear at grazing in both these azimuths. As the bombarding potential is increased the beams move up from the surface to satisfy the relation $\lambda = d \sin \theta$. Ten or fifteen degrees above the surface radiation from the second and lower layers escapes in sufficient amounts to reduce the intensity of the resultant beam through interference, and at a somewhat higher angle the beam disappears.

An exactly similar beam is found at grazing in the C-azimuth. The grating constant here is 1.24 Å. and the bombarding potential corresponding to wave-length 1.24 Å. is 97.5 volts. The beam appears at grazing at just this voltage. These three beams occurring and behaving exactly as required by the theory constitute the strongest evidence we have in favor of the wave interpretation of electron scattering.

We have been less successful in trying to account for the occurrences of the remaining 21 sets of beams. We do not know why they occur

where they do. The most we have been able to do is to relate their occurrences with those of the Laue beams that would issue from the same crystal if the incident beam were a beam of x-rays.

In Fig. 7 we indicate by crossed circles in a (λ, sin θ) diagram the x-ray diffraction beams that would be observed in the B-azimuth. We show also again the electron beams as actually observed. It is obvious that the law of occurrence of electron beams is not the same as the law of occurrence of Laue beams, and yet we see that the occurrences of the two sets of beams have certain features in common. The dots representing electron beams occur along the plane grating lines at about the same intervals as the crossed circles representing the Laue beams. Other points of similarity are found with further study of the data and one is led finally to the conviction that each electron beam is the analogue of a particular Laue beam. The electron beam represented by a given dot appears to be the analogue of the Laue beam of the same order represented by the crossed circle occurring next above it in the diagram. This association of beams is indicated in the figure.

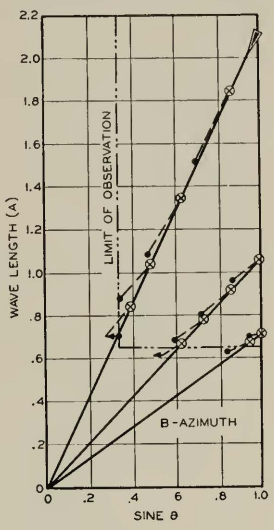


Fig. 7—λ sin θ diagram for B-azimuth

The occurrences of the Laue beams are determined in part by the separation between the atomic plane gratings that make up the crystal. If the separation between adjacent planes were increased the crossed circles representing the Laue beams would be moved upward along the plane grating lines; if the separation were decreased the crossed circles would be moved downward. Merely as a mode of description, then, we may say that a given electron beam has the wave-length and position that its Laue beam analogue would have if the separation between planes were decreased by a certain factor.

We have calculated this spacing factor for each of the 21 beams and the values found are plotted in the upper part of Fig. 8 against the voltages of the beams. The points form a very bad curve. They do indicate, however, that the factor increases with the speed of the electron, and there is the suggestion that it approaches unity as a limiting value. There is the suggestion, that is, that at high voltages the law of occurrence of electron beams is the same as the law of occurrence of Laue beams.

It has been pointed out by Eckart that if the index of refraction of the crystal for the electron radiation is other than unity diffraction beams will occur as if the separation between atom planes were other than normal. We have computed the indices of refraction that would

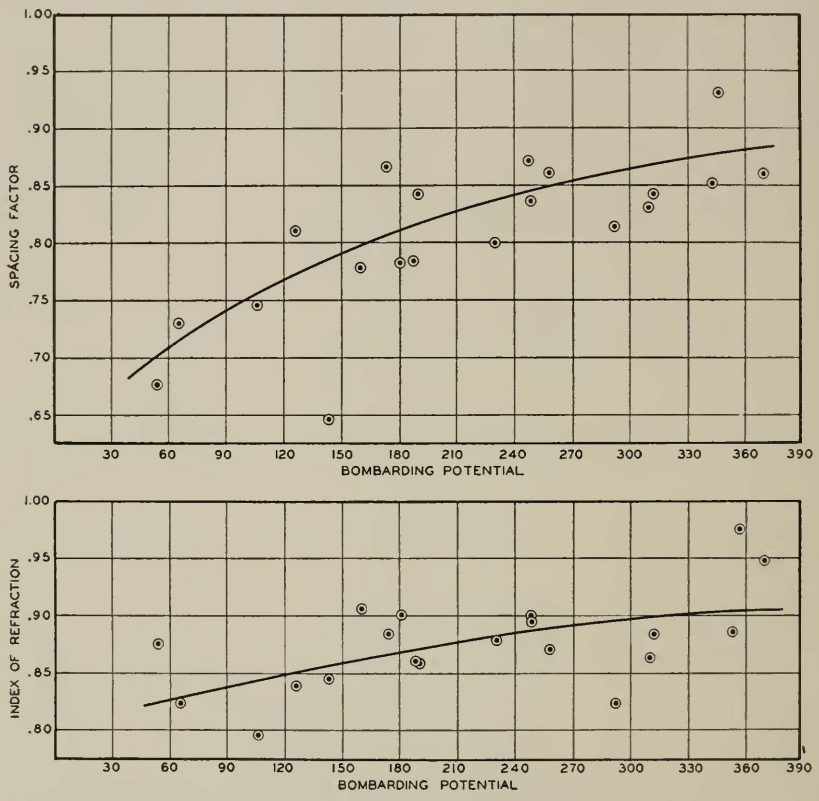


Fig. 8—Plot of values of spacing factor and associated values of refractive index for twenty-one beams

give rise to the observed occurrence of beams and these are plotted in the lower part of the diagram against bombarding potential. Again the points fall very irregularly. While it cannot be said that there is at present a satisfactory explanation of the peculiar occurrence of the space lattice electron diffraction beams, it should be clearly understood that this deficiency in no way affects either the wave-length measurements of these beams or the agreement of these wave-lengths with the values of h/mv.

The electron diffraction beams which I have described are the only ones observed when the surface of the crystal is free from gas. When the surface is not free from gas still other beams appear. These

beams are due to the scattering of electrons by the adsorbed gas and therefore we shall not consider them at this time.

In closing I should like to say a few words about the conceptual difficulty in which these experiments involve us. When Laue and his collaborators investigated the scattering of x-rays by crystals the results of their observations were accepted at once as establishing the wave theory of x-rays. It was a very simple matter for W. H. Bragg and others to give up the corpuscular theory because of the hypothetical nature of the x-ray corpuscle. It was only necessary to recognize that Laue's results were contrary to hypothesis and the corpuscle disappeared.

If the electron were not the well-authenticated particle we know it to be, it is possible that the experiment I have described would cause it to vanish in like manner. We do not, however, anticipate any such event. The electron as a particle is too well established to be discredited by a few experiments with a nickel crystal. The most we are apt to allow is that there are circumstances in which it is more convenient to regard electrons as waves than as particles. We will allow perhaps that electrons have a dual nature—when they produce tracks in a C. T. R. Wilson cloud experiment they are particles, but when they are scattered by a crystal they are waves.

A quite similar situation exists, of course, in the case of x-rays. It has been evident for some years that the adherents of the corpuscular theory of x-rays were too enthusiastic in their recantations. X-rays also exhibit a dual nature—when they give rise to diffraction patterns they are waves, but when they exhibit the Compton effect or cause the emission of electrons from atoms they are particles—quanta or photons.

This state of affairs is one that should appeal to us as intolerable. There must, it would seem, be comprehensive modes of description applicable to all electron and x-ray phenomena, but what these are we do not yet know. We do not know whether we shall eventually believe with de Broglie and Schroedinger that electrons and x-rays are waves that sometimes masquerade as particles, with Duane that electrons and x-rays are particles that sometimes masquerade as waves, or whether eventually we shall believe with Born that we are dealing in both cases with actual particles and phantom waves.

I believe, however, that for the present and for a long time to come we shall, in describing experiments, worry but little about ultimate realities and logical consistency. We will describe each phenomenon in whatever terms we find most convenient.

Grid Current Modulation

By EUGENE PETERSON and CLYDE R. KEITH

Synopsis: The term grid current modulator is used to describe those vacuum tube circuits in which modulation is initially produced in the grid circuit of a three-electrode vacuum tube due to the non-linear grid current-grid voltage relation. Comparison with a representative plate current modulator using the same tubes and the same plate potential shows that by modulating at maximum efficiency in the grid circuit and using the plate circuit solely for amplification, the maximum power output is increased about eight times, the power efficiency is increased about five times and the ratio of sideband output to signal input is increased approximately three times. Under these conditions more carrier input power is needed for the grid than for the plate modulator. This improved performance has been made possible by a detailed study of the fundamental processes involved and by a design of the tubes and associated equipment, such as transformers and filters, to permit these fundamental processes to operate to their best advantage.

Normally modulation is also produced in the plate circuit which is shown to be out of phase with that produced in the grid circuit. By inserting high impedances to the input frequencies in the plate circuit, plate circuit modulation is prevented, and the reduction of grid circuit sideband is likewise avoided. By including in the grid circuit an impedance which is high to the desired sideband frequencies, the maximum grid sideband voltage is obtained. In this way the power and modulating efficiencies of the tube circuit are made maximum.

Where modulation occurs only in the plate circuit of a tube, the sideband amplitude is proportional to the product of the amplitudes of the input frequencies when these amplitudes are small. In the present type of grid current modulator the sideband amplitude is proportional to the smaller of the two input amplitudes provided the ratio between these is greater than about 3/2. This feature makes the modulator particularly valuable in communication systems.

The article concludes with an application of the fundamental principles involved to an experimental carrier telephone system in which the operating features of tubes, filters, and transformers are discussed.

Introduction

BECAUSE of the extensive application of vacuum tube modulators in systems of carrier communication, they constitute an important tool in the hands of telephone engineers. As such they have justified extensive laboratory investigation. The purpose of this paper is to discuss some of the properties of a type of modulator utilizing the non-linear relation existing between grid voltage and grid current and the advantages which recent laboratory investigations indicate that it may possess. Further studies are in progress to determine the conditions under which it can be employed practically.

There are two distinct classes of vacuum tube modulators which may be designated for convenience as grid and plate types, according to the circuit in which the modulation is initially produced, although some modulators may involve both circuits. As an example of the plate

type we might mention the coupled-plate or Heising modulator which has found extensive application in radio transmitters, in which the plate circuit of an oscillating tube is coupled to the plate of another tube through which the signal is introduced, modulation taking place ordinarily in the plate circuit of the oscillating tube. A carrier frequency amplifier is sometimes used in place of the oscillator. Another type of plate modulator, due in principle to van der Bijl, which has found extensive application in carrier telephone systems, applies the signal and carrier to the grid of the modulator, so that the two components are amplified in common before being modulated. As examples of the grid modulator there are the grid leak and condenser type used almost universally for radio reception, together with the type which forms the subject of the present paper, employing a generalized impedance in the grid circuit. The three last-mentioned modulators may incidentally produce modulation in both plate and grid circuits. This ordinarily acts to reduce the overall efficiency as well as to introduce other undesirable features of operation, so that in the design of the grid current modulator we have been led to minimize modulation in the plate circuit, operating the grid circuit as a modulator and the plate circuit purely as an amplifier.

The criteria of usefulness of modulators include some usually placed upon vacuum tube apparatus in general, together with those peculiar to frequency change; some of most importance are modulating gain and level, plate power efficiency, quality, stability, input and output impedances, and carrier suppression. These will be taken as a basis for discussing the operation of the grid current modulator and comparing it with that of the other types mentioned above. The modulating gain usually expressed in transmission units (T. U.) represents the ratio of the power output of a single sideband to the power input of the signal which produces it. It is a function of both carrier and signal amplitudes, usually decreasing at high amplitudes. This decrease in gain should not be too rapid or the modulated output power at sufficiently large signal amplitudes may actually decrease as the signal is increased, and lead to prohibitive distortion. Another aspect of the question relates to the maximum modulated power attainable. The signal amplitude fluctuates within wide limits in course of operation and it becomes desirable to limit its effects, so that the resultant modulated potentials may not disturb the operation of associated equipment. This may be accomplished in modulators in which the sideband output approaches an asymptotic maximum as the signal is increased, better than in those which pass through a maximum in the operating range. A knowledge of the signal amplitude and the

modulating gain corresponding to it is sufficient for a determination of the output amplitude or output level which is of prime importance in matters relating to noise and interference. Other significant factors from the standpoint of noise and interference are the closeness with which line and connecting apparatus impedances are matched since this determines the amount of reflection of an incident wave,[1] and the extent to which carrier current is transmitted to the line (carrier leak) in carrier suppression systems.

It is highly desirable of course to have the efficiency of energy conversion from the plate battery to the sideband output power as high as possible, since the amount of power supplied to the plate circuit is thereby minimized, and the necessary power capacity of the plate supply is reduced. Another kind of plate efficiency in which we are sometimes interested is the efficiency of energy transfer from the plate supply to the external plate impedance. This tells us the amount of energy dissipated by the plate of the tube and fixes its structure; it differs from the first efficiency only when other current components than the sideband flow in the plate circuit. Inasmuch as we shall deal in the following with low power tubes which have ample load capacity, only the first-mentioned efficiency is important.

We are also interested ordinarily in the quality of the system. This is determined in large part by the width of the transmitted frequency band, by the presence of new interfering frequencies introduced by the process of modulation, and by the linearity of the modulated output in terms of the signal input. The first and last conditions are equivalent to the requirements that the modulating gain be maintained both over the ordinary range of signal input amplitudes, and over the frequency band essential for good signal reproduction. Finally the system as a whole is required to have a high degree of stability, so that ordinary variations of battery potentials, or even the replacement of a tube by another of the same type, will not impair the operation of the system.

The specific forms of grid current modulator with which we shall be concerned as an application of the theory are those adapted to carrier current telephony, in which the carrier current is suppressed and a single sideband transmitted. Comparison with a representative plate current modulator using the same tubes at the same plate potential shows that, by modulating at maximum efficiency in the grid circuit and using the plate circuit solely for amplification, the maximum power output is increased eight times, the power efficiency is increased five times, and the ratio of sideband output to signal input is increased

[1] The reflection is measured by the quotient of the difference by the sum of the two connected impedances; this is known as the reflection coefficient.

approximately three times. From these figures it is evident that the space current or plate power supplied the grid modulator is sixty per cent greater than it is for the plate modulator, and that this greater power is utilized five times more efficiently. Under these conditions more carrier input power is needed for the grid than for the plate modulator. Where the carrier oscillator employs a tube of the same type as that used in the modulators, sufficient carrier power is, however, available. This improved performance has been made possible by a detailed study of the fundamental processes involved, and by a design of the tubes and associated equipment, such as transformers and filters, to permit these fundamental processes to operate to best advantage.

In view then of the close interdependence of the circuit elements, we shall start with a discussion of the theory as developed for the simplest circuits, and accompany it by approximate mathematical analyses wherever it appears profitable. No rigorous mathematical treatment appears to be possible or even desirable because of its complexity in the general case, and the sole purpose of our approximate analyses is to help in building up a physical picture of the operation of modulators. With the theoretical conclusions in mind, the characteristics of tubes, transformers, retard coils, balanced circuits, and filter networks which are important in this connection are examined, and the performance of the complete carrier telephone modulator circuit is covered in some detail. The theoretical conclusions are not limited to the carrier telephone modulator which has been used simply for illustrative purposes; as a matter of fact the same principles have been found operative in different types of circuits over a wide frequency range. It should be noted that we have not attempted to combine the oscillating and modulating functions in a single circuit as is sometimes done, but have maintained these circuits distinct from one another, so that the best performance of each may be realized.

Theory of Vacuum Tube Modulation

In a broad sense the same general phenomena are involved in both grid and plate modulation, since modulation is produced when an impedance is varied in accordance with the amplitudes of the modulating potentials, a condition true of both circuits under appropriate conditions. Thus conductive grid current is suppressed at negative potentials in the high vacuum tubes which we employ, and it flows when the grid potential is positive, the grid impedance depending upon the amplitude of the grid potential.

We can obtain a qualitative idea of the situation when we consider

the grid circuit connected to an a.c. generator in series with a high resistance of the order of a megohm, and with the plate circuit connected to a resistance of the same order of magnitude as the internal plate resistance. With a negative grid the input impedance is mainly capacitive and at low frequencies nearly the entire applied e.m.f. exists across the grid. At positive potentials however, when conductive grid current flows and the grid resistance drops to something like 10,000 ohms, by far the greater part of the drop is taken up in the external resistance so that the positive lobe of the grid potential wave is distorted. This distortion is equivalent to modulation since it implies the presence of new frequencies. As a result of the varying reaction of the tube then, modulation voltages are built up across the grid-filament path. These potentials are amplified in the plate circuit where the applied wave suffers further distortion due to the non-linear relation between plate current and grid potential, which in a similar way gives rise to plate modulation. Evidently plate modulation would be alone effective if the external grid resistance were made small.

General Relations

The production of modulated currents or potentials is characteristic of any device in which the relation between instantaneous values of current and voltage is not a linear one. Theoretically, such a relation can be represented to any required degree of accuracy by an equation of the form

$$i = a_0 + a_1 v + a_2 v^2 + a_3 v^3 + \cdots, \tag{1}$$

in which i represents the current through the device and v the potential drop across it. Now suppose a voltage wave which includes two components of frequency $p/2\pi$ and $q/2\pi$ respectively to be impressed on the non-linear element

$$v = P \cos pt + Q \cos qt. \tag{2}$$

If we substitute this expression in eq. 1 for v, the current wave is found to include components of the two original or fundamental frequencies, together with new frequencies produced by the non-linear element:

$$\begin{aligned} i = i_0 &+ i_p \cos pt + i_q \cos qt \\ &+ i_{2p} \cos 2pt + i_{2q} \cos 2qt \\ &+ i_+ \cos (p + q)t + i_- \cos (p - q)t \\ &+ i_{3p} \cos 3pt + i_{3q} \cos 3qt \\ &+ i_{2p+q} \cos (2p + q)t + i_{2p-q} \cos (2p - q)t \\ &+ i_{p+2q} \cos (p + 2q)t + i_{p-2q} \cos (p - 2q)t + \cdots, \end{aligned} \tag{3}$$

in which the coefficients i_k involve the characteristic constants of the tube, together with the applied potential amplitudes:

$$\begin{aligned} i_0 &= a_0 + a_2(P^2 + Q^2)/2 + \cdots, \\ i_p &= a_1P + 3a_3P(P^2 + 2Q^2)/4 + \cdots, \\ i_q &= a_1Q + 3a_3Q(Q^2 + 2P^2)/4 + \cdots, \\ i_{2p} &= a_2P^2/2 + \cdots, \\ i_{2q} &= a_2Q^2/2 + \cdots, \\ i_+ &= i_- = a_2PQ + \cdots, \\ &\vdots \end{aligned} \tag{4}$$

The new frequencies produced are made up of sums and differences of integral multiples of the two original frequencies, and an inspection of eq. 3 shows that the frequency of any component may be put in the form

$$|mp \pm nq|/2\pi \qquad m, n = 0, 1, 2.\cdots$$

It is convenient to designate the sum of the two numbers m and n as the order of a wave component, so that the frequencies $2p/2\pi$, $2q/2\pi$, and $(p \pm q)/2\pi$ are products of the second order. The last of these serves as the basis for the operation of all present [2] carrier systems, and the rôle of any modulator is therefore to produce one or both of these components which are known as side frequencies, or as sidebands when the signal wave is made up of a band of frequencies. Now by repeating the modulating process, but this time with the frequencies $(p + q)/2\pi$ or $(p - q)/2\pi$, or both, together with the component of frequency $p/2\pi$, designated as the carrier wave, it is well known that one of the resultant second order products has the frequency of the original signal $q/2\pi$. This second or receiving modulator, sometimes designated as a demodulator or detector, is separated from the first one by a transmitting medium and frequency-selective apparatus so that only the desired components may be transmitted and received. The impedance-frequency characteristics of these elements with which the modulators are associated are of prime importance in determining the modulation, as is best brought out by a discussion of some approximate mathematical analyses which follow.

Modulator Circuit, Small Alternating Potentials

We shall consider the current-voltage characteristic of a vacuum tube to be given, to sufficient accuracy for our purposes, by the first

[2] Higher order products, such as $(2p \pm q)/2\pi$ have been equally well employed but we shall confine our attention here to the usual second order system.

three terms of eq. 1, and shall suppose two generated potentials, of frequency $p/2\pi$ and $q/2\pi$ respectively, to be applied to a tube circuit which includes a series impedance. This impedance Z_k may be a function of frequency as indicated by the subscript k, which refers to the particular frequency at which the impedance is effective. The variable part of eq. 1—the change in current produced by application of the alternating potentials—is clearly

$$J = a_1 v + a_2 v^2. \tag{5}$$

The potential drop across the tube is that impressed minus the Zi drop or

$$v = \Sigma(E_k - Z_k J_k) \tag{6}$$

the summation extending over all current components. As a first approximation to the fundamental currents we may neglect the non-linear term ($a_2 v^2$) in eq. 5 to obtain

$$\begin{aligned} J_p &= a_1 E_p/(1 + a_1 Z_p), \\ J_q &= a_1 E_q/(1 + a_1 Z_q). \end{aligned} \tag{7}$$

Using these solutions we can obtain a second approximation [3] taking into account the non-linear term which we neglected for the first approximation. Thus

$$\begin{aligned} J_p + J_q + J_2 = a_1 \left(\frac{E_p}{1 + a_1 Z_p} + \frac{E_q}{1 + a_1 Z_q} - Z_2 J_2 \right) \\ + a_2 \left(\frac{E_p}{1 + a_1 Z_p} + \frac{E_q}{1 + a_1 Z_q} \right)^2, \end{aligned} \tag{8}$$

in which the subscript 2 indicates the second approximation. By squaring the second member of eq. 8 it is observed that the second approximation includes direct current, second harmonics of the two impressed frequencies $p/2\pi$ and $q/2\pi$, and the second order sidebands. For these last we obtain the expression

$$J_{\pm} = \frac{a_2}{(1 + a_1 Z_{\pm})} \frac{E_p E_q}{(1 + a_1 Z_p)(1 + a_1 Z_q)}. \tag{9}$$

The sideband potential across the variable element is clearly $Z_+ J_+$ or $Z_- J_-$ according to the particular sideband in which we are interested. Thus

$$V_- = -\frac{a_2}{a_1^2} J_p J_q \frac{Z_-}{1 + a_1 Z_-} = -\frac{a_2}{a_1^2} J_p J_q \frac{Z_-}{R_0 + Z_-}, \tag{10}$$

where $1/a_1$ is equivalent to R_0, the plate resistance of the tube.

[3] Carson, "A Theoretical Study of the Three Element Vacuum Tube," *Proc. I. R. E.*, 1919.

These equations are equally applicable to grid and to plate circuits provided the potentials are small and the operating region is expressible by eq. 5. This is not always the case in practice, and modifications in the above comparatively simple analysis are required, which will be treated below. A number of characteristic features of operation are exhibited by the above analysis however and these we proceed to discuss.

If the preceding treatment is applied to the grid circuit we see that in order to make the sideband potential across the grid a maximum with fixed fundamental currents, the external grid impedance at the sideband frequency must be made large compared to the effective internal resistance of the tube. Further, if the generator potentials and impedances are fixed it follows that the generator resistance should be made to match the internal resistance of the tube at the fundamental frequency—with a transformer, if necessary—in order to make the fundamental currents as large as possible. This conclusion regarding the ratio of grid impedances follows immediately without mathematical analysis if we suppose the source of the higher order products to lie in the variable impedance element so that it may be considered equivalent to the presence of generators of the higher order frequencies. The generator voltage is evidently maximum on open circuit, which agrees with the above statement.

In considering the form of external impedance to use for best results, an inspection of eq. 10 shows that in the quantity $Z_-/(R_0 + Z_-)$, which expresses the ratio of effective grid voltage to generated grid voltage, the sideband impedance Z_- should have a large reactive component. This is illustrated by Fig. 1 which gives the ratio for various relative external impedances having phase angles of 0 and 90° and shows that, with the external impedance fixed in magnitude, the ratio has its greatest value for a pure reactance.

Relative Phase of Grid and Plate Sidebands

If the tube acted as a perfect amplifier of the potentials impressed on the grid, there would be no further distortion and the sideband current in the plate circuit would be obtained by multiplying eq. 10 by $\mu/(Z + R_0)$, where Z and R_0 are the external and internal plate circuit resistances respectively, and μ is the amplification factor which is assumed constant here.[4] Unfortunately this ideal situation does not exist of itself, and modulation of the amplified fundamentals takes place in the plate circuit, producing an additional sideband component to

[4] The distortion due to variable μ as treated by Peterson and Evans in the *Bell System Technical Journal* for July, 1927, represents but a small part of the total in efficient modulators, although it is of importance in high quality amplifiers.

combine with that generated in the grid circuit and amplified in the plate circuit.

Inasmuch as the amplification factor decreases, and the plate impedance increases as the grid potential goes negative, the grid

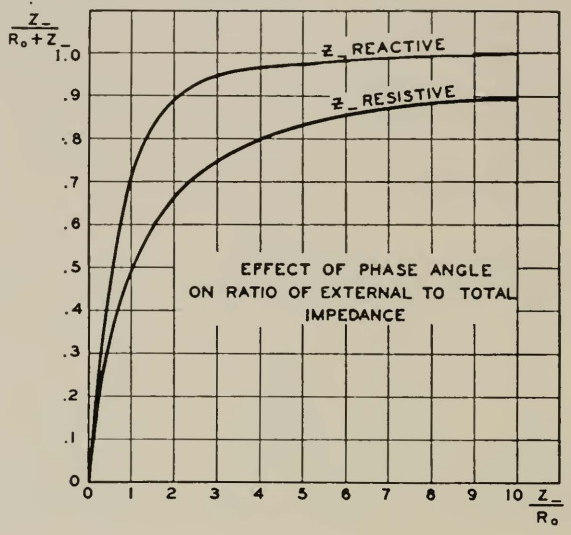


Fig. 1

potential wave is amplified more efficiently on the positive than on the negative lobe, with the result that the plate current wave is limited on one side by the grid current cut-off, and on the other side by plate current cut-off. The second cut-off tends to make the output wave more nearly symmetrical about a horizontal axis; it is therefore equivalent to an increase in the odd order modulation which we do not employ here, and to a reduction of the even order products, one of which—the second order sideband—is used to transmit signal characteristics. It follows that for efficient modulation we must do one of two things—phase the grid and plate products to add, or remove one of the conflicting sources of modulation.

To account for the effect of plate distortion we may apply the same general procedure to the plate circuit as we did to the grid circuit. The plate current-grid potential relation is given as

$$J = b_1\mu v + b_2\mu^2 v^2 \qquad (11)$$

and the solution for the current components may be written down directly since the problem presents itself in the same form as the grid circuit situation previously considered. Hence if we change the a

coefficients to b's, and the Z_k of the grid circuit to the Z_k of the plate circuit, we obtain the expressions

$$\left.\begin{aligned} J_1 &= b_1\mu v/(1 + b_1Z_1), \\ J_2 &= \frac{b_2}{(1 + b_1Z_2)}\left(\frac{\mu v}{1 + b_1Z_1}\right)^2. \end{aligned}\right\} \tag{12}$$

Now it is apparent that each of these two terms contributes something to the sideband frequency—the first by amplification of the grid sideband potential, and the second by modulation of the two fundamental components in the plate circuit. The net sideband current in the plate circuit may accordingly be expressed as

$$J_\pm = \left[\left(\frac{\mu b_2}{(1 + b_1Z_p)(1 + b_1Z_q)} - \frac{b_1a_2Z_\pm}{1 + a_1Z_\pm}\right)\right] \left[\frac{\mu E_pE_q}{(1 + b_1Z_\pm)(1 + a_1Z_p)(1 + a_1Z_q)}\right]. \tag{13}$$

Under normal conditions both grid current and plate current characteristic curves are concave upward, so that b_1, b_2, and a_2 are all positive. The two terms of eq. 13 are then in phase opposition, a condition which is responsible for failure to work certain modulators to fullest advantage. For efficient plate modulators the grid modulation term should be suppressed, which may be accomplished by making the external grid impedance to the modulated product of interest equal to zero, or by keeping the grid potential negative at all times. For efficient grid modulators the plate modulation term should be reduced to a minimum by suppressing the fundamental currents in the plate circuit, in which case Z_p and Z_q are made large. Of course the possibility exists of phasing the two sideband components to add rather than to subtract (arithmetically)—and this, it will be readily seen, is obtained by having the phase angle of the entire plate circuit approach 90° at each fundamental frequency when the grid circuit sideband impedance is large. This condition cannot be met without lowering the amount of plate current modulation, so that the first mentioned plate circuit condition is the more practical one.

Other possibilities of more favorable phasing exist by working within appropriate regions of tube operating characteristics where either b_2 or a_2 becomes negative. Generally speaking, operating points of this nature are not stable with variations in tube potentials, nor are they adaptable to large power outputs approaching the maximum load capacity of the tube. Finally, for straight amplification purposes the two terms of eq. 13 should be made equal and opposite in sign.

In order to test directly the conclusions regarding relative phase of grid and plate modulation products, a circuit was set up which permitted two frequencies to be supplied to the grid of a vacuum tube, the resultant currents of sideband frequency being measured in grid and plate circuits by means of a current analyzer.[5] As shown in Fig. 2 the

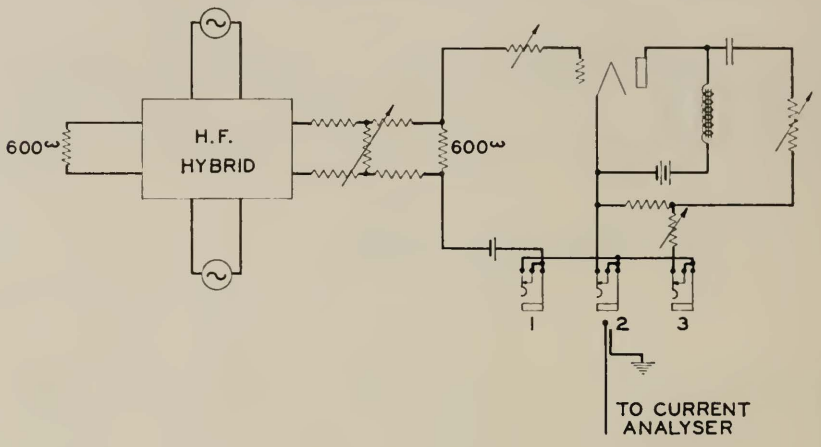


Fig. 2. Test Circuit

grid circuit contains a high external resistance (for producing grid current modulation when conductive grid current flows as previously explained) in series with a "C" battery to vary the relative amounts of grid and of plate modulation. The relative phase of sideband currents produced in grid and plate circuits was calculated from the currents measured separately in jacks No. 1 and No. 3 and their vector sum in jack No. 2. These measurements verified the conclusions drawn from eq. 13,—that with resistances in grid and plate circuits second order modulation products produced in the grid circuit are exactly out of phase with the same frequencies produced in the plate circuit.

The effect of the grid circuit resistance when conductive current flows is of course to limit the positive potentials applied to the grid and so, in effect, to cause the input-voltage—output-current relation of the circuit to be deflected at the upper end more nearly to parallelism with the x-axis than it is for the tube alone. We may therefore consider grid modulation as equivalent to the introduction of a reversed curvature in the operating characteristic. To substantiate this point a tungsten filament tube was used in which the curvature of the lower branch is nearly the same as that of the upper branch, as shown in Fig.

[5] "Analyzer for Complex Electric Waves," by A. G. Landeen, *Bell System Technical Journal*, April, 1927.

3. The plate circuit sideband was measured as a function of the "C" battery with zero grid resistance so that the plate circuit was responsible for the total sideband production; the data are plotted on the

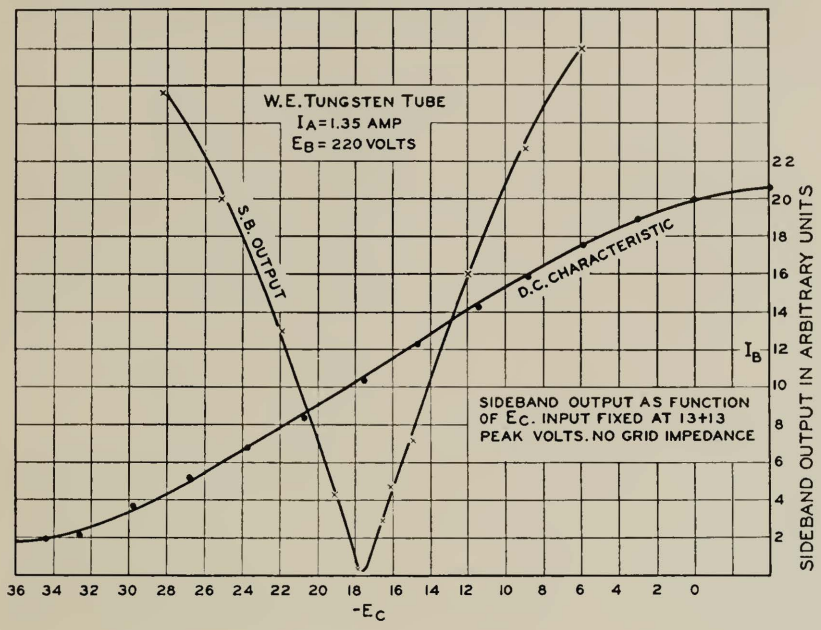


Fig. 3

same Figure. It is seen that the sideband drops nearly to zero when the "C" battery is adjusted so that the input voltage swings symmetrically over the upper and lower branches of the curve. These results could very well be attributed to out-of-phase modulation resulting from reversed curvature. As a matter of fact the algebraic expression for the characteristic involves in general a series of both odd and even powers of applied voltage, but if the axis is taken at the point of symmetry of the characteristic the even powers drop out. Now since the even orders of modulation can be attributed only to the even powers of the static equation, it might be expected that these components would drop to zero.

We may conclude from this discussion that best results will be had in the practical design of grid current modulators, when the external grid impedance at the sideband frequency is made as high as possible and when the impedance to the fundamentals is matched. As to the plate impedances, the situation is the reverse of that existing in the grid circuit since we must have the impedances to the two modulating

frequencies as high as possible, and match the tube impedance at sideband frequencies in order to develop maximum sideband power in the load resistance. It will be observed that with these conditions satisfied, the plate circuit of the tube acts substantially as an amplifier of the sideband produced in the grid circuit, since none is developed in the plate circuit.

Reaction of Sideband Flow on Tube Impedance

In any non-linear system such as the grid circuit or the plate circuit of a tube, the modulation products resulting from the lack of linearity have amplitudes which depend upon one or more of the impressed fundamentals, and react upon the fundamental amplitudes. It follows that the amplitude of any modulation product depends in general upon the amplitudes of all other modulation products, and that the impedance offered to the flow of fundamental depends upon the reaction of the modulation products. Stated otherwise, the amplitude of any one current component depends upon all other components.

This may be demonstrated quantitatively by higher approximations than the two which we have already obtained, in which expressions for the currents are found to contain terms proportional to the sideband voltage across the tube; in fact, if we went to the labor of including a number of distorting components, terms in the fundamental current equation due to their reaction would result. The effects found with a single sideband are simply typical. If, for example, we put (7) and (9) in (8), we get

$$J_p = a_1(E_p - Z_p J_p) - a_2(E_q - Z_q J_q)\,\frac{a_2 E_p E_q Z_+}{(1 + a_1 Z_p)(1 + a_1 Z_q)(1 + a_1 Z_+)},$$

and a similar expression for J_q, as second approximations to the fundamentals. These furnish us with a pair of simultaneous cubics in J_p and J_q. When we assume the reaction of the sideband flow on J_q to be small so that eq. 7 remains valid, the above equation becomes linear in J_p,

$$J_p = E_p \Big/ \left(\frac{1}{a_1} + Z_p + \frac{a_2^2}{a_1^4} J_q^2 \frac{Z_+}{1 + a_1 Z_+}\right). \qquad (13a)$$

This shows that the impedance in the fundamental path has been increased due to non-linearity by the amount of the last term which may be denoted by ΔZ_p where

$$\Delta Z_p = \frac{a_2^2}{a_1^4} \frac{Z_+}{1 + a_1 Z_+} J_q^2.$$

A similar expression exists for ΔZ_q when J_p is substituted for J_q. The reciprocal of a_1 will be recognized as the internal resistance of the variable element for small potential variations.

From these

$$J_p^2 \Delta Z_p = J_q^2 \Delta Z_q,$$

so that the two fundamental circuits share equally in the power dissipation due to sideband flow. This means that when the two modulating currents are not of the same amplitude, the smaller current will have the larger resistance change due to sideband flow, and therefore will suffer a greater percentage amplitude change. This discussion serves to emphasize the point that the tube impedances depend upon the impedance-frequency characteristics of the circuit to which the tube is connected, so that this point must be kept in mind in the design and measurement of modulating circuits.

Grid Current Modulator, Large Alternating Grid Potentials

The comparatively simple analysis we have just employed is not capable of very wide application because of the assumed form of the grid current equation. In the practical forms of grid current modulators, from which comparatively large amounts of modulated power are required, the grid potentials are increased and the grid is maintained negative during an appreciable part of the cycle. The above method then becomes too involved to be extended to this case, since a large number of terms would be required for an accurate representation of the tube characteristic. When we have a large external grid resistance, however, as appeared to be desirable from eq. 10, a fairly exact solution for the modulation products can be obtained by another method which is capable of direct application.

If we determine the relation between impressed potential and output current in this particular case we find that on passing from negative to positive potentials, the plate current curve breaks sharply at about zero grid potential, and becomes nearly parallel to the x-axis, as shown in Fig. 4. We can therefore consider the positive lobe of the input wave to be cut off at zero grid potential under these conditions and the problem can be handled analytically.[6]

We are indebted to Mr. F. Mohr for computations on the sideband amplitude as given by eq. 20, of the Appendix, in which the sideband is expressed in terms of a multiple of P, as function of the ratio Q/P. The relationship between these quantities is given as a single-valued function. For our own purposes, however, we have plotted the

[6] Appendix 1.

sideband potential as a function of one of the modulating potentials with the other as parameter, as shown in the dotted lines of Fig. 5. The experimental data are plotted as the full lined curves and appear

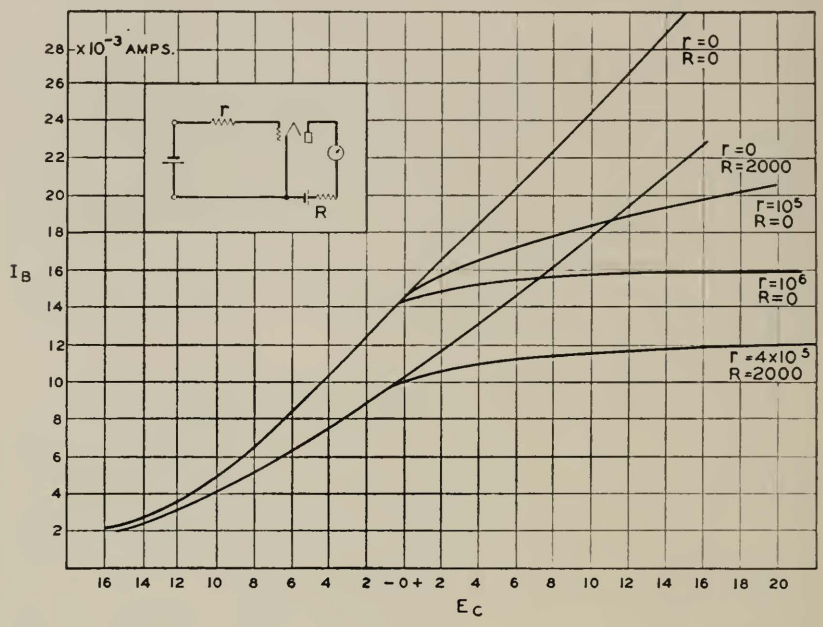


Fig. 4

to be in good agreement with the theory, the divergence being due presumably to the incomplete suppression of the positive lobe in the experimental set-up. The measurements were carried out with the current analyzer as described in connection with Fig. 3, grid current being measured as a function of the two input amplitudes. The sideband grid potential was then determined by multiplying the sideband current by the external grid resistance.

Possibly the most striking thing shown by Fig. 5 is that the sideband amplitude is independent of the larger of the two inputs, when the ratio of one input to the other is made sufficiently great. Hence we must provide sufficient carrier amplitude to insure that the resultant sideband shall be linearly proportional to the impressed signal up to its greatest value so that good quality of speech transmission may be assured.

Our earlier analysis using eq. 5 to represent the grid current characteristic led to a sideband amplitude proportional to the product of the two inputs whereas in this case in which the positive lobe is completely

suppressed, it is proportional to the smaller of the two when the ratio of the two inputs is greater than about 3/2. This proceeds from the fact that eq. 5 must be supplemented by many more terms involving higher powers of the impressed potentials, in order to represent the

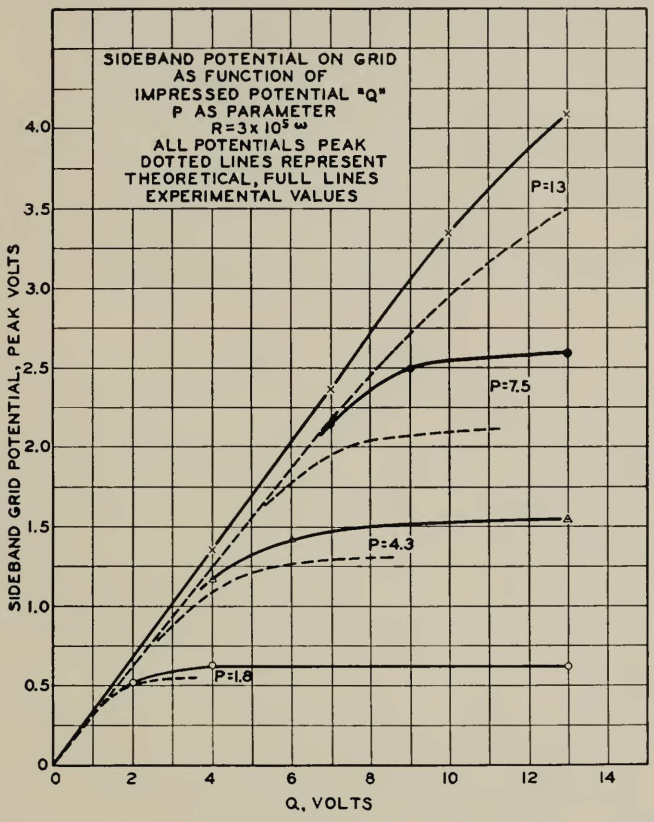


Fig. 5

grid current characteristic to any degree of precision when the grid is driven negative.

The form of the input-output curve is especially valuable for telephony. The relative independence of the larger of the two inputs means that the sideband output will be stable with regard to carrier current variations under the limitations noted. The output approaches a maximum asymptotically, so that the articulation at heavy loads may be expected to hold up better than in those modulating systems in which the output passes through a pronounced maximum. In a system transmitting the carrier, as in radio, and in which a square law

detector is used, the voice output is proportional to the product of the received carrier and sideband. Any change in attenuation, expressed in transmission units (T. U.), between the transmitting and receiving station affects the output current by twice that number of T. U. In the above grid current demodulator, however, the output changes to the extent of the attenuation change, and varies no more than in a carrier suppression system with the carrier locally supplied.

Having determined the grid voltage components, we may now apply the plate circuit coefficients to the grid potential in order to determine the plate current components, just as we did in the previous case. There, it will be recalled, we used a simple representation for the plate current in terms of the grid potential from which amplification and modulation terms were deduced. The same general considerations regarding phase opposition are carried over unchanged.

Limitation of Sideband Output

The above method of treatment is quite satisfactory when the space current is never reduced to zero, but when the grid voltage goes sufficiently negative, precisely the same limitations apply to the plate characteristic equation as applied to the grid equation under similar circumstances, and there exists an additional source of distortion in the plate circuit. In this circumstance the method of expansion in Bessel coefficients cannot readily be used because of the large number of components in the wave subjected to additional distortion, which would lead to prohibitive complication. We may nevertheless obtain a qualitative idea of the result in special cases of interest to us in this connection.

We shall assume that, as we found previously to be advisable, one of the two fundamental currents is substantially suppressed in the plate circuit, so that despite the non-linearity of the plate circuit sideband components are produced only in the grid circuit. We are therefore concerned with the variation of amplification with operating parameters. Now it is clear to start with, that at sufficiently small grid potentials, the entire variation falls within the region of variation of the plate dynamic characteristic so that the result may be written down as in the previous analysis. As the amplitude is increased, the negative end of the grid swing finally has no effect in varying the space current, and the distortion which results tends to limit the magnitude of the amplified components. Hence as the sideband potential on the grid is increased by increasing the applied signal potential and keeping the carrier potential large enough (say one and one half times the signal) so as to get the full efficiency of grid current modulation, the

sideband current in the plate circuit increases linearly with the signal up to a certain point. At this point, which corresponds to the plate current cut-off, the output departs from the linear relation and increases less rapidly. Further increase of the carrier produces no increase in output but a reduction of the output may result because of a greater swing beyond the cut-off point.

Inasmuch as the modulating potentials together with undesired modulated products form a wave having a net amplitude considerably greater than that of the useful sideband, it is clear that the maximum output amplitude can be increased by suppressing the undesired current components thus avoiding the loading and heating effects produced in large part by these other components. A method of attaining this desired result will be treated in connection with balanced circuits. The loading effect may be partially ameliorated very simply since one of the products of modulation is a d.c. component. The presence of series grid resistance means that we have in effect a negative bias applied to the grid which becomes increasingly negative as the input amplitude increases,—just the sort of thing, in other words, to limit sideband production. If, therefore, we use grid reactances instead of grid resistances we can achieve the same degree of modulating efficiency in the two cases at low inputs, and in addition remove effective grid bias, the maximum output power available being increased to a very considerable extent. Of course the insertion of grid reactance changes the details of the conclusions for the grid resistance, but the main features of performance are retained.

When grid resistance is used to provide a high impedance to the sideband, the operation of the grid leak and condenser detector is approached, in respect to the undesirable increase of bias with increase of input. As a consequence the output power is limited at large inputs, although the gain is fairly high at small input amplitudes.

Another point affecting the operation of the grid leak and condenser detector is the plate circuit impedance. According to the conclusions of the above theory for grid current modulation, the output power is increased at large input amplitudes by providing an impedance in the plate circuit which is high to both input frequencies and matches the tube impedance at all desired output frequencies. This conclusion has been verified experimentally at carrier frequencies when operating the tube for maximum output, but is contrary to the usual practice in radio circuits, where the plate circuit impedance to the modulating frequencies is ordinarily made low rather than high compared to the tube impedance. The problem is complicated at radio frequencies by regenerative effects not present to the same degree at the compara-

tively low frequencies used in carrier telephony, and by the comparatively low alternating and battery potentials which raise the relationship of plate and grid voltages to grid current, to importance.

We have now to examine the electrical properties of available circuit elements in the light of our previous analysis, so that their assembly will yield the most favorable results.

Vacuum Tubes

The effect of the shape of the grid-current—grid-voltage curve on the modulating properties of the grid circuit is not as pronounced at large amplitudes as might be expected from experience with plate current modulators at comparatively low amplitudes. As is well known this characteristic of ordinary tubes is much more variable between tubes of the same type than the plate-current—plate-voltage curve. But it has been found that a change of tubes having static grid characteristics varying within wide limits does not vary the modulating gain of a grid current modulator more than one T.U. The reason for this may be seen most easily in the case of an external grid impedance consisting of a pure resistance. If the tube grid resistance were comparatively small for all positive voltages the positive half of the wave would be completely suppressed, and the analysis of Appendix 1 would accurately represent the wave. Even when the tube grid resistance varies considerably it does not alter the wave shape appreciably so long as it remains small compared to the external resistance. This condition may be satisfied with particular ease for large input voltages, and may also be satisfied in a qualitative sense, when reactances are used in place of resistance. The principal effect of a change in grid resistance is then to change the input impedance, which affects the net gain only through the mismatch of impedance at input frequencies.

As a consequence of the tube circuits and range of operating potentials used in the grid current modulator, the details of the grid current characteristics become of relatively small importance and attention is focussed on the functioning of the plate circuit. The plate circuit is used purely for amplification purposes as mentioned above, so that the criteria of usefulness of a tube as a grid current modulator come down ordinarily under the stated operating conditions to the criteria of usefulness of a tube as an amplifier.

Filter and Transformer Networks

Input Filters and Modulating Gain

Since the gain obtainable in a grid current modulator depends primarily on the ratio of external to total grid circuit impedance, it is

necessary to consider how the required high impedance may be obtained in practice. The input transformer must have an impedance looking into the grid side which is high to all sideband frequencies, and must at the same time transmit efficiently all signal input frequencies. A high impedance over the sideband range is best obtained by a filter [7] on the low side of the input coil, care being taken to allow for the effect of the transformer on the filter impedance. In order to determine the actual external grid impedance and to investigate the modification of filter impedance by the input transformer, a high impedance bridge was built in which precautions were taken to prevent errors due to the high impedances involved (up to several megohms). In each case only the end section of the filter adjacent to the modulator was used, since this provided nearly the same impedance as would be given by a complete filter. Low pass filters are used on the input to the modulator and on the output of the demodulator, while band pass filters are used on the output of the modulator and the input of the demodulator. These are to be considered in turn.

Low Pass Filters

The simplest type of low pass filter is the infinity type of section, the impedance characteristics of which are shown in Fig. 6*a*. The filter alone, as shown by the solid lines, has negligible reactance in the transmission band (0–3 K. C.) and practically pure inductance in the attenuated region. The input transformer resonates in the attenuated region when terminated by this filter, as shown by the dotted lines, because of the leakage inductance and distributed capacity of the windings. The resonance peak is quite broad due to the comparatively high a.c. resistance and so covers a considerable frequency range as is shown by the ratio of $Z_-/(R_0 + Z_-)$ in Fig. 6*c*. It may be made to appear at higher frequencies by using a filter with a higher cutoff frequency, and at a lower frequency by replacing the series inductance by a parallel tuned circuit (an *m*-type section).[8] A new type of filter section developed for certain phases of this work and known as the built-out type [9] has a particularly good impedance characteristic in the attenuated region, as shown by the solid lines of Fig. 6*b*. But as shown by the dotted lines the transformer impedance, when terminated in this type of section, is very much modified by the coil constants. The resulting efficiency as show by the ratio $Z_-/(R_0 + Z_-)$ in Fig. 6*c* is not as good as that of the infinity type section.

[7] For a general discussion of filter impedances and attenuations see Campbell, *Bell System Technical Journal*, November, 1922; Zobel, January, 1923; Johnson and Shea, January, 1925.

[8] O. J. Zobel, *Bell System Technical Journal*, October, 1924.

[9] Devised by T. E. Shea of Bell Telephone Laboratories.

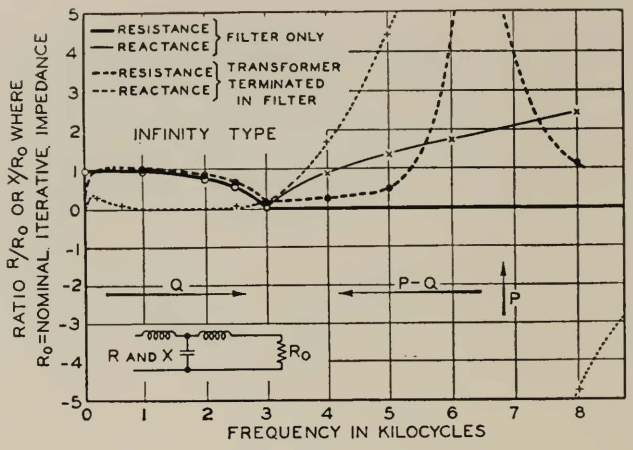


Fig. 6a

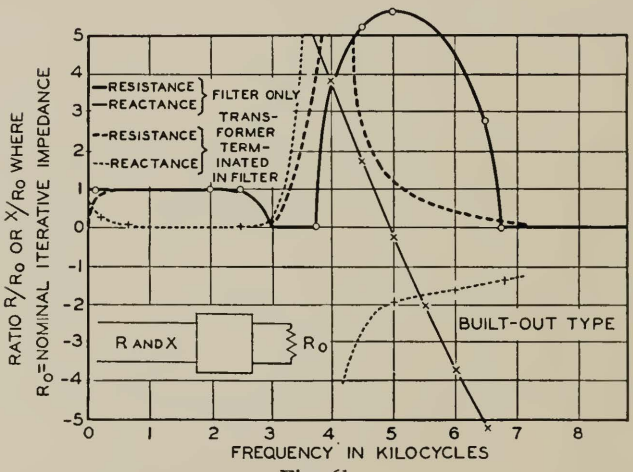


Fig. 6b

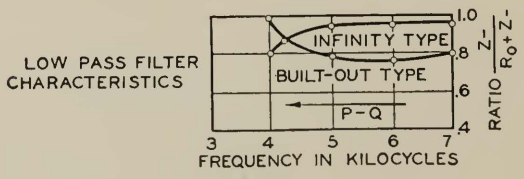


Fig. 6c

Band Pass Filters

The two most important types of band pass filter sections—the confluent and the built-out types—are shown in Fig. 7. The resonance of the confluent section in the voice frequency range does not affect the

ratio of external to total impedance appreciably but the chief defect is in the comparatively low value of this ratio at the higher voice frequencies (2.5 to 2.8 K. C.), although this type of section is satisfactory over the entire voice frequency range at higher carrier frequencies.

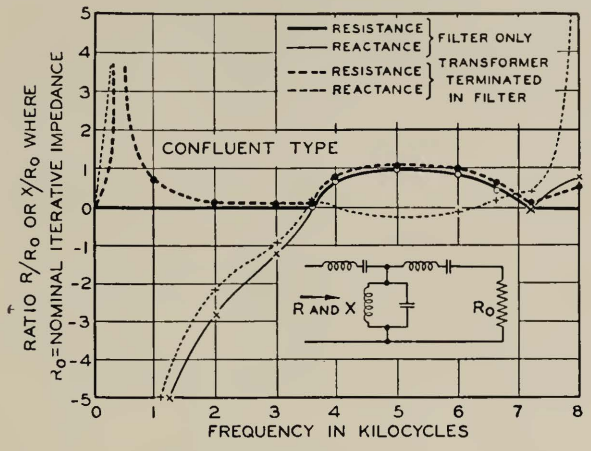


Fig. 7*a*

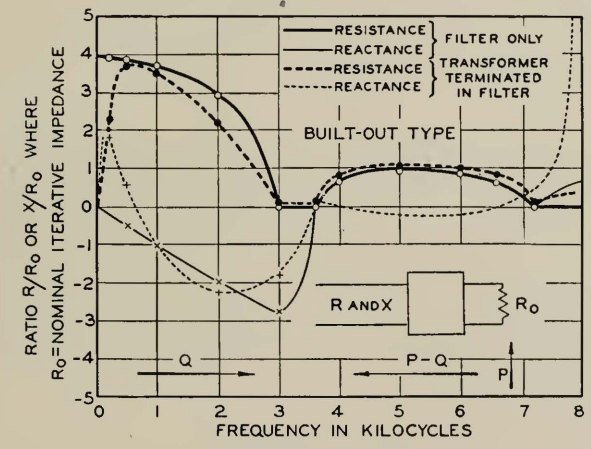
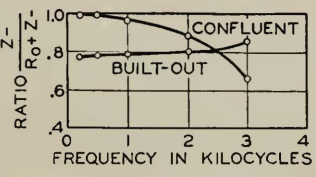


Fig. 7*b*



BAND PASS FILTER CHARACTERISTICS

Fig. 7*c*

The built-out band pass filter shown in Fig. 7*b* has a very satisfactory impedance over the voice range and the modifications introduced by the transformer do not seriously affect its efficiency. From the curves of Fig. 7*c* it is evident that the built-out type of section must be used for channels near the voice frequency range but that the confluent type shown in Fig. 7*a* may be used for the higher frequency channels.

The close relation between input filter impedance and modulating gain is illustrated in Fig. 8. Two band pass filters were built having

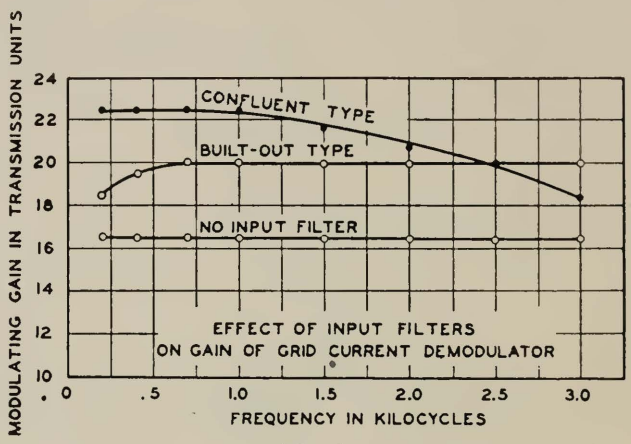


Fig. 8

impedance characteristics approximating the curves shown in Figs. 7*a* and 7*b*. It would then be expected that the modulating gain would be proportional to the ratio $Z_-/(R_0 + Z_-)$. The curves in Fig. 8 show that this is very nearly the case. With no input filter the ratio $Z_-/(R_0 + Z_-)$ would be 0.5 or 6 T.U. less than the maximum possible gain, as is found to be actually the case. This shows that the modulating gain may be calculated for any value of input impedance if it is known for any other value.

Input Impedance

The impedance looking into the low side of the input transformer when the high side is terminated in the grid circuit of a modulator under operating conditions (carrier at normal value) depends on a number of factors, among which the principal variables are the signal and carrier input currents, the input transformer, and the input generator impedance. As might be expected, the input impedance decreases as either carrier or signal amplitude is increased, and the change of impedance with signal amplitude is small when the signal is small compared to the carrier, as is normally the case.

The influence of the input transformer upon input impedance depends not only upon first order, but also upon higher order effects. The first order effect is simply due to the transformer terminated in a network having a linear current-voltage characteristic, which may be calculated from the usual transformer theory. The higher order effect is produced by the effect of the contributions to fundamental frequencies caused by the flow of modulation currents, as discussed in connection with Equations 13*a* and 4. For this reason the impedance of the external grid circuit at other than input frequencies may have a considerable effect on the input impedance. It has been found possible to reduce the reflection from a resistance line to a small value with suitable transformers.

Output Filters and Transformers

The general effect of an output filter or retard coil in the plate circuit with high impedance to all frequencies except the sideband is to increase the output level for large inputs, since the opposing effect of plate modulation is eliminated and the total load capacity of the tube is employed solely in the amplification of the sideband. The output transformer on account of its low ratio has very little effect in altering the impedance-frequency characteristic of the output filter so that we need not enter so thoroughly into the details as we did in the case of input filters.

Output Impedance

The output impedance of a grid current modulator (looking from the line into the output coil) is affected mostly by the transformer ratio and the impedance to carrier in the plate circuit. If the impedance to the carrier frequency is very high, as is usually the case, there will be very little modulation with the carrier in the plate circuit, and neither the carrier input current nor the external output impedance at signal frequencies affects the output impedance appreciably. The reflection may be made quite small over the frequency range without any great difficulty.

Gain-Frequency Characteristic

The problem of obtaining a flat frequency-gain characteristic over the voice range depends upon the attenuation of input and output transformers, the attenuation of filters, and the impedance characteristic of input and output coils when terminated by their respective filters. The transformer attenuation is comparatively small and affects the frequency characteristic mostly at frequencies below 200 cycles. The closer the carrier channels are spaced to each other or to the voice band, the more difficult it becomes to obtain filters with suf-

ficiently sharp cutoff. In most cases a maximum variation of 2 T.U. in the attenuation over the transmitted band is a reasonable figure for a band pass filter. Each transformer and filter tends to increase the attenuation at the edges of the transmitted band more than in the center so that frequencies from 800 to 2,000 cycles are always transmitted with minimum attenuation, which is independent of the frequency-output current characteristic of the modulating elements.

The above consideration of filter attenuation is substantially independent of filter impedance since the latter is determined mostly by the end section. From the previous consideration it is evident that either may have a very pronounced effect, so that in measuring the frequency characteristic of a modulator or demodulator both attenuation and impedance effects must be taken into account. The effects of input and output impedances can be partially separated when the carrier is suppressed because of the fact that the output impedance has but little effect at small inputs and the input impedance has but little effect at large input currents.

Balanced Tube Circuits

The present practice in carrier telephone systems is to suppress the carrier current and one sideband in order to conserve frequency space and to reduce the energy levels and the cross-talk in associated equipment. The elimination of undesired components of a wave may be carried out by two distinct processes,—frequency discrimination by filter networks, and phase discrimination or balance by bridge circuits.[10] Each method is useful and both find places in carrier systems. When the frequency separation between desired and undesired components becomes relatively small, frequency discrimination becomes impractical and expensive. The balance method is used to separate frequencies according to their respective phase relations in two or more similar modulating circuits, the phases of the output components depending on the relative phases of the input currents. Consequently only certain combinations of modulation products can be separated by balance and these only to an extent determined by the balance attainable in transformers and vacuum tubes, both of which are subject to manufacturing variations. Due to the proximity of the carrier and second order sideband frequencies the suppression of carrier current by filter circuits alone is impractical. Balanced circuits must be used for this purpose and in spite of unavoidable variations in tubes and circuits it is usually possible to reduce the carrier on the line to less than five per cent

[10] For an illustration of balanced circuits, reference may be made to U. S. Patent 1,343,306, issued to J. R. Carson.

of its normal unbalanced value. To separate one sideband from the other after the carrier has been suppressed, and to suppress unbalanced components other than the carrier, filter attenuation is customarily employed.

In the usual type of balanced circuit there are two possible input paths with corresponding output circuits, one connected to the two grids in series; and known as the series path; the other to the two grids in parallel, known as the shunt path or midbranch. When carrier is impressed on the midbranch and signal on the series arm—the present arrangement in commercial carrier systems using plate current modulators—we designate the circuit, as a matter of convenience, as the "Conjugate Input Type." The modulation product frequencies are distributed as shown in Fig. 9*a*. When both signal and carrier are impressed on the series branch the modulation product frequencies are

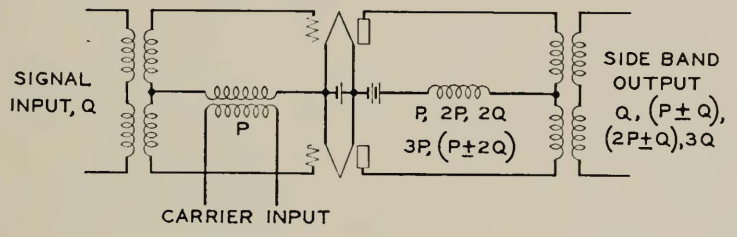


Fig. 9*a*. Conjugate Input Type

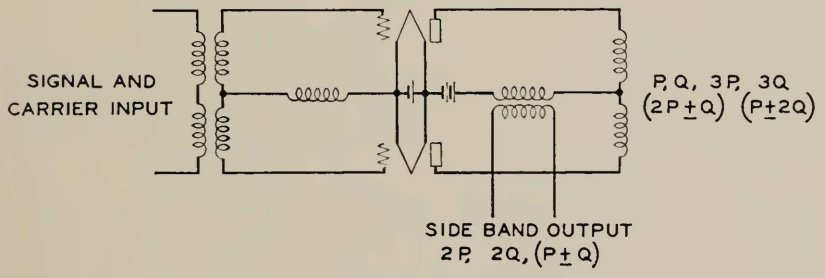


Fig. 9*b*. Common Input Type

as shown in Fig. 9*b* and the circuit is called for convenience the "Common Input Type." The phase of the modulation product of any order may be determined from the consideration that the phase of the frequency

$$\frac{1}{2\pi}|mp \pm nq|$$

depends upon the quantity $(m\theta_p \pm n\theta_q)$ where θ represents the phase

angle between current and voltage of each input frequency. This conclusion is independent of the type of modulation employed. If the phase of the product is then calculated to be identical on the two grids or two plates, it appears in the midbranch; if it turns out to be opposite in phase on the two grids referred to the filament, it appears in the series arm.

Conjugate Input Grid Modulator

With the signal introduced in the series arm of Fig. 9*a*, the sideband potential is built up across the same arm, so that a high sideband impedance must be provided by the input transformer terminated in its filter,—a low pass filter for the modulator, and a band pass filter for the demodulator. The carrier frequency is introduced in the conjugate arm so that the carrier circuit does not directly affect the signal and sideband impedances. There is a second order effect, however, due to the reaction of those modulation products which flow in the common branch. The input impedance may be expected to change also when the coupling between the two high impedance windings of the input transformer is varied, since this effectively changes the impedance to the above mentioned modulation products.

Modulation is largely eliminated in the plate circuit and the load capacity of the tubes is increased by inserting a choke coil in the common plate branch to suppress the carrier current. This, incidentally, tends to reduce carrier leak. The impedance of the choke coil at the carrier frequency is modified by the capacity to ground of the output transformer, and must be designed with this point in mind since the shunt arm impedance may otherwise be materially reduced. Some further increase in load capacity is obtained by having the output transformer and the terminating filter offer a high impedance to frequencies outside the transmitted band. In this way all important components except the sideband are suppressed and the plate circuit of the tube operates as an amplifier so that the plate power dissipation is reduced and the load capacity increased. The same considerations regarding the second order impedance effects of the shunt branch on the series branch exist for the plate circuit as for the grid circuit considered above. The main effect when there is loose coupling between the high impedance windings of the output transformer is to introduce an inductive reactance into the series arm. This tends to increase the reflection coefficient so that it becomes preferable to couple the two windings closely, a comparatively easy thing to do in low impedance circuits. The modulation products accompanying the desired product are indicated in Fig. 9, and it is seen that there will be no introduced distortion up to the third order when the carrier frequency is sufficiently high.

Common Input Grid Modulator

Another useful type of grid current modulator is shown in Fig. 9*b*, in which both signal and carrier are applied across the same input terminals. The modulation currents flow in the plate (and corresponding grid) circuits as shown in the above schematic. Where the ratio of carrier to signal frequency is large so that a single input transformer cannot be used efficiently, separate transformers with associated filter networks may be used for each of the two inputs. Since the second order sidebands ($p \pm q$) appear in the midbranches, it is not necessary to have the impedance high to these frequencies in the input coil, but only from the midpoint of the input coil to ground. This is most conveniently accomplished by a high inductance retard coil in the midbranch of the grid circuit, although transformers and high impedance networks may be used in general. The grid circuit sideband across the midbranch is amplified and appears in the plate circuit midbranch. The fundamental currents together with all odd order modulation products are eliminated by a high impedance, high mutual retard coil in the series arm of the plate circuit.

Since the present practice is to use suppressed carrier, a hybrid [11] coil must be used to introduce the carrier if this circuit is to be used as a demodulator, although the signal and carrier currents may be introduced through filters when used as a modulator. Either frequency discrimination or balance is required in any case to keep carrier current out of the signal circuit.

The chief advantage of the common over the conjugate input type of circuit is that the high impedance required for the modulated product is provided by a distinct element, and no high impedance requirements are placed on other elements in either input or output circuits. Another advantage of this arrangement is that the amplified fundamentals are balanced out, making the singing gain about 20 T.U. less than that of the conjugate input type. The only modulation products (up to the fourth order) not balanced out of the output are the second harmonics of carrier and signal. This type of circuit may be used as a demodulator at any frequency, but as a modulator only when the second harmonic of the highest voice frequency does not come in the sideband range—it is therefore not well adapted to modulate low carrier frequencies where high quality is required.

Although the output of this modulator is affected but little by the filter impedance in either input or output circuits, some care is neces-

[11] By using a hybrid coil having eight times as many turns in the signal circuit as in the carrier circuit, the equivalent current losses to signal and carrier are 0.5 T.U. and 9.5 T.U. respectively instead of 3 T.U. each, as is the case for the usual equality ratio hybrid coil.

sary in selecting the retard coils for the grid and plate circuits. Since the grid retard should have a high impedance to the desired modulation frequencies it must have an inductance of the order of 50 henries or greater at low frequencies in a demodulator. Resonance in the voice band is not harmful so long as the impedance does not drop too much at high voice frequencies.

The plate circuit retard coil is well balanced to reduce the unbalanced carrier transmitted to the line. An important requirement is that of close coupling so that the reactance in the output circuit may not be great enough to cause a transmission loss or large reflection coefficient. The required inductance then depends upon the relative separation of voice and sideband frequencies. If the lowest sideband frequency is very close to the highest voice frequency it may be impossible to prevent positive reactance from coming into the voice circuit of a demodulator, but the effect may be considerably reduced by utilizing this positive reactance in the mid-series section of the adjacent low pass filter.

Double Balanced Circuits

If two balanced circuits of either of the above types are connected with their input and output terminals respectively in series, all the modulation products up to the fourth order except the second order sidebands may be balanced out. There is no hybrid or filter loss and due to more complete suppression of unwanted frequencies the maximum output power obtainable is more than twice that with a single balanced circuit. The complexity of the resultant circuit is such as to rule it out for all ordinary applications.

For purposes of comparison we proceed to consider the experimental results obtained on a conjugate input grid modulator designed in accordance with the ideas set forth above.

Experimental Results

Figs. 10 and 11 represent the results of experiment on a conjugate input grid modulator with a carrier frequency of 6,800 cycles and a signal frequency of 1,000 cycles. The input and output networks previously discussed and represented in Figs. 6 and 7 were used here with 101-D tubes operated at 120 volts plate potential and 1.0 ampere filament current. The grids were connected to the negative terminal of the filaments. Fig. 10 represents the sideband output current in a 675 ohm circuit, plotted as a function of the signal current measured in the 675 ohm input circuit, with the carrier input maintained at 15 mils throughout. The upper four curves represent various experimental conditions designed to bring out the effect of different circuit

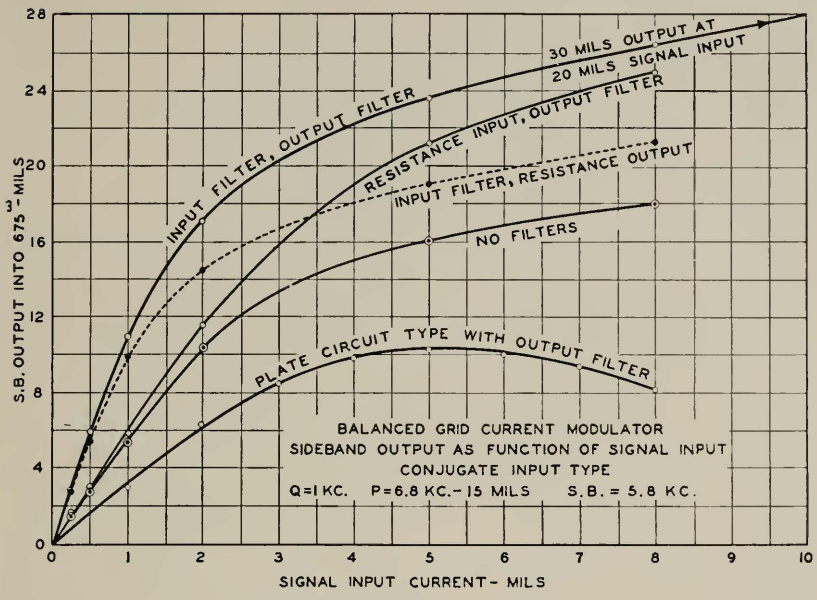

30 MILS OUTPUT AT
20 MILS SIGNAL INPUT
INPUT FILTER, OUTPUT FILTER
RESISTANCE INPUT, OUTPUT FILTER
INPUT FILTER, RESISTANCE OUTPUT
NO FILTERS
PLATE CIRCUIT TYPE WITH OUTPUT FILTER
BALANCED GRID CURRENT MODULATOR
SIDEBAND OUTPUT AS FUNCTION OF SIGNAL INPUT
CONJUGATE INPUT TYPE
Q=1 KC. P=6.8 KC.-15 MILS S.B. = 5.8 KC.
S.B. OUTPUT INTO 675 Ω - MILS
SIGNAL INPUT CURRENT - MILS

Fig. 10

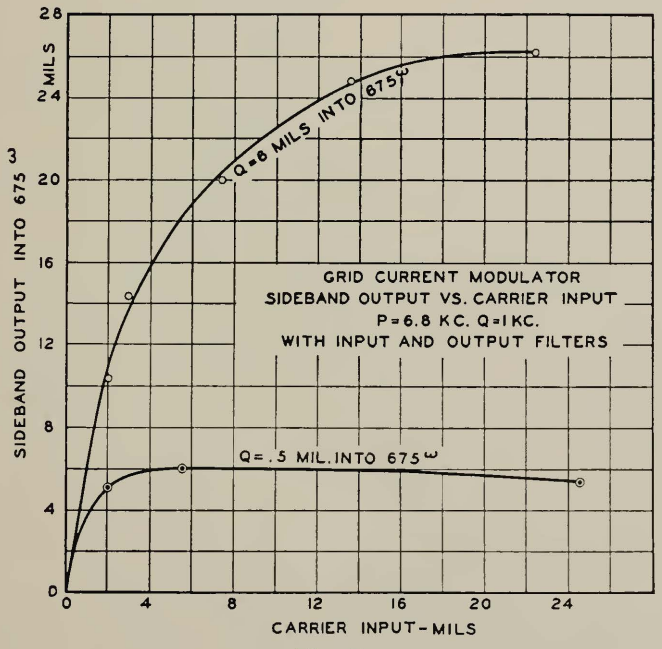

Q=6 MILS INTO 675 Ω
GRID CURRENT MODULATOR
SIDEBAND OUTPUT VS. CARRIER INPUT
P=6.8 KC. Q=1 KC.
WITH INPUT AND OUTPUT FILTERS
Q=.5 MIL. INTO 675 Ω
SIDEBAND OUTPUT INTO 675 Ω - MILS
CARRIER INPUT - MILS

Fig. 11

elements, while the lowest curve illustrates the performance of a representative conjugate input plate modulator working under conditions prescribed for it into a 600 ohm circuit with the same tubes and plate potential. A direct comparison between the two types of modulator as to sideband current output should include a comparative increase of 0.6 T.U. to the grid current modulator output to take care of the difference in the two load impedances.

The curve labelled "no filters" applies to the circuit of Fig. 9 in which both input and output circuits were connected to 675 ohm resistances. The presence of the retard coil in the plate circuit is accountable for the increase in output at large signal inputs over that of the plate type. When an output filter is added (resistance input, output filter) the gain at low inputs is scarcely affected but the output power for large signals is doubled since the load capacity is increased by the suppression of the signal frequency current in the plate circuit. If now an input filter is inserted and the output connected to a 675 ohm circuit (input filter, resistance output) the gain at low signal currents is increased by about 5 T.U. over that with no filters in circuit, while the increase at high signal amplitudes is of the order of 1.5 T.U. The topmost curve represents the performance of the modulator circuit terminated in the two filters, which shows a modulating gain of 21.5 T.U. at small inputs and a maximum power output of 30 mils into 675 ohms (0.6 watt). Fig. 11 represents the effect of varying the carrier input at two signal inputs—0.5 and 6 mils respectively. This illustrates the lack of dependence of sideband on carrier when the carrier is greater than the signal, which was deduced from eq. 20 as characteristic of this type of modulator. The use of a 15 mil carrier is seen to furnish close to the optimum value for the circuit, at least when the signal amplitude does not greatly exceed 6 mils.

The common input type is capable of yielding much the same results as the conjugate input type with somewhat less care required for the flanking filter impedances, since the proper circuit impedances are obtained by the use of retard coils as shown in Fig. 9*b*. On theoretical grounds, however, as we mentioned in discussing the general properties of balanced circuits, it is not capable of furnishing as high quality as the conjugate type at the low sideband frequency used here. At high sideband frequencies this objection disappears, so that the reduced filter requirements make it perhaps more attractive in application than the conjugate type. It should be noted that the plate modulator may be made to have a greater gain than that shown in Fig. 10 by changing the input transformer (with the same maximum output level) but this restricts the signal input current to correspondingly smaller amplitudes.

A few words on the shape of the signal-sideband curve of plate modulators of the van der Bijl type may not be inappropriate at this point since the curve depends to some extent upon the incidental grid modulation produced. Thus at large signal amplitudes the grid of the modulator tube is driven positive and grid modulation is produced, which tends to oppose plate modulation. By reversing the conditions which we have employed in the grid current modulator to promote grid modulation, the net plate modulation may be increased and the sideband-signal curve may more nearly show an asymptotic maximum which is so desirable from the overloading standpoint. This condition is evidently secured with a flanking input filter having a low impedance to the sideband, or by having an input coil which, while not seriously affecting the transmission of signal frequencies, offers of itself a low impedance to the grid sideband. Thus in plate modulators the input coil would have a high winding capacity, and in plate demodulators it would have comparatively low mutual inductance between primary and secondary windings.

As an indication of the quality obtained with the grid current modulating process, comparative listening tests between carrier telephone systems employing plate and grid demodulators, respectively, conducted by R. W. Chesnut, indicate roughly a 10 T.U. greater load carrying capacity for the grid type over a wide range of input amplitudes at about the same quality in both cases. The carrier leak may be reduced to one half mil by a not very critical tube selection, which is quite satisfactory in general.

The last point remaining is the plate power efficiency, which we have defined as the ratio of the sideband power developed in the load resistance to the d.c. power supplied to the plate circuit under operating conditions—it is really the efficiency of power conversion. At maximum output it is three per cent for the standard plate modulator and fifteen per cent for the above grid modulator. The efficiencies obtained at maximum output for a number of different low power tubes used in the grid current modulator may be tabulated as follows:

Tube	Plate Potential	Sideband Power W_{AC}	Plate Efficiency W_{AC}/W_{DC}
230-D.....	60	0.022	11%
221-A.....	70	0.065	18
221-D.....	90	0.13	14
101-D.....	120	0.50	15
102-D.....	120	0.11	22

For design information and construction of the experimental models,

the authors are indebted to E. B. Payne and H. R. Kimball for filters, and to H. Whittle and A. G. Ganz for transformers and retard coils.

Appendix

Grid Current Modulator, Large Grid Potentials

Making use of the observation that the positive lobes of the input wave are effectively suppressed with a sufficiently large external grid resistance, we first define a function equal to zero when the independent variable is positive, and equal to the variable when the variable is negative. This is evidently a representation of the potential effective on the grid in terms of the applied potential. If we denote the grid potential by $-f(y)$ where y is the impressed potential, it may be expressed as a Fourier series

$$-f(y) = b_0/2 + \Sigma b_m \cos m\pi y/Y + a_m \sin m\pi y/Y, \tag{14}$$

in which the coefficients are determined by the usual relations

$$\left.\begin{aligned} b_m &= \frac{1}{Y}\int_0^Y y \cos\frac{m\pi y}{Y}\,dy = \frac{Y}{m^2\pi^2}(\cos m\pi - 1), \\ a_m &= \frac{1}{Y}\int_0^Y y \sin\frac{m\pi y}{Y}\,dy = (-1)^{m-1}\frac{Y}{m\pi}\cos m\pi, \\ b_0 &= \frac{1}{Y}\int_0^Y y\,dy = Y/2. \end{aligned}\right\} \tag{15}$$

In these equations y represents the generator e.m.f. and Y is its maximum value. If we put (15) in (14), we have

$$-f(y) = Y/4 - \frac{2Y}{\pi^2}\sum_{m=1}^{m=\infty}\frac{\cos(2m-1)\pi y/Y}{(2m-1)^2} + \frac{Y}{\pi}\sum_{m=1}^{m=\infty}\frac{(-1)^{m+1}}{m}\sin\frac{m\pi y}{Y}, \tag{16}$$

the last term of which may be summed to $y/2$ and (16) may be rewritten as

$$-f(y) = Y/4 + y/2 - \frac{2Y}{\pi^2}\sum_{m=1}^{m=\infty}\frac{\cos(2m-1)\pi y/Y}{(2m-1)^2}, \tag{17}$$

which represents the desired solution. It will be observed that the first two terms of the right member contribute only a d.c. term and the fundamentals, so that the other modulation products must come from the summation. This expression is a perfectly general one as far as the form of y is concerned. In the case of a sinusoidal grid e.m.f. it is possible by more customary methods to find the grid potential, but

with a complex grid potential in which we are primarily interested no simpler representation is known to the authors except where the two frequencies involved are harmonically related.

With the two frequency inputs considered, we have the grid potential

$$y = P \cos pt + Q \cos qt \tag{18}$$

and the summation term may be written

$$\frac{2(P+Q)}{\pi^2} \sum_{m=1}^{m=\infty} \frac{\cos\left[(2m-1)\pi(P \cos pt + Q \cos qt)/(P+Q)\right]}{(2m-1)^2}. \tag{19}$$

Upon expansion of (19) as the cosine of the sum of two angles we find the terms cos $(A \cos \theta)$ and sin $(A \cos \theta)$ which require evaluation before the solution can be put in significant form. This might conceivably be done by direct expansion; for example, the first of the expressions would then be

$$\cos (A \cos \theta) = 1 - \frac{(A \cos \theta)^2}{2!} + \frac{(A \cos \theta)^4}{4!} \cdots.$$

Putting the terms of this series in terms of multiple angles gives us an infinite series to be summed as the coefficient of each multiple angle term. This may be done in terms of Bessel coefficients by Jacobi's expansions [12] which are as follows

$$\cos (A \cos \theta) = \sum_{n=0}^{n=\infty} (-1)^n \epsilon_{2n} J_{2n}(A) \cos 2n\theta,$$

$$\sin (A \cos \theta) = \sum_{n=0}^{n=\infty} (-1)^n \epsilon_{2n+1} J_{2n+1}(A) \cos (2n+1)\theta,$$

in which $J_k(A)$ is a Bessel coefficient of the kth order and ϵ_k is Neumann's factor which is two for k not zero, and unity for k zero. Carrying out the expansion, the sideband amplitude comes out as

$$\frac{4(P+Q)}{\pi^2}\left[J_1\left(\frac{P\pi}{P+Q}\right) J_1\left(\frac{Q\pi}{P+Q}\right) + \frac{1}{3^2} J_1\left(\frac{3P\pi}{P+Q}\right) J_1\left(\frac{3Q\pi}{P+Q}\right) + \cdots\right], \tag{20}$$

which may be computed from tables to be found in Watson's book [13] previously cited. Analogous expressions exist for the other components.

[12] Watson, "Theory of Bessel Functions," p. 22.
[13] P. 666 et seq.

A High Efficiency Receiver for a Horn-Type Loud Speaker of Large Power Capacity

By E. C. WENTE and A. L. THURAS

Synopsis: This paper describes a telephone receiver of the moving coil type which is particularly adaptable to the horn type of loud speaker and which represents a notable advance over similar devices at present available. Its design is such as to permit of a continuous electrical input of 30 watts as contrasted with the largest capacity heretofore available of about 5 watts. In addition, measurements show that the receiver has a conversion efficiency from electrical to sound energy varying between 10 and 50 per cent in the frequency range of 60 to 7,500 cycles. Throughout most of this range, its efficiency is 50 per cent or better. This contrasts with an average efficiency of about 1 per cent for other loud speakers either of the horn or cone type. Combining the 50 fold increase in efficiency with a 5 or 6 fold increase in power capacity, a single loud speaker unit of the type here described is capable of 250 to 300 times the sound output of anything heretofore available.

This device is in commercial use in connection with the Vitaphone and Movietone types of talking motion pictures. As commercially produced in quantities numbering several thousand, an average efficiency of the order of 30 per cent has been realized.

BEFORE the advent of radio-broadcasting, practically the only loud-speakers in commercial use were of the horn type. In recent years this type has been largely supplanted by others of more compact design. However, where appearance and size are not of prime importance, a loud speaker with a horn still has a large field of service, as, for instance, in public address equipment or in systems for reproducing speech and music in large auditoriums from wax or film records. For such uses, the greater directivity obtained by the use of a horn has in some respects definite advantages. In the design of the receiver about to be described we have had in view particularly the requirements for such services, where the following qualifications were deemed of the greatest importance: a good response-frequency characteristic up to at least 5,000 p.p.s., large power output without amplitude distortion, high efficiency, and constancy of performance.

As this paper is concerned with the design of a driving unit, or the receiver proper, and not with the horn, we shall confine our discussion to the operation of the receiver when connected to a tube of infinite length and of the same cross-sectional area as the throat of the horn. An ideal horn should have at its throat the same acoustic impedance [1]

[1] The term acoustic impedance as here used may be defined as the ratio of pressure to rate of volume displacement.

as a tube of this character, viz., $\frac{\rho c}{A}$ c.g.s. units, where ρ is the density of air, c, the velocity of sound, and A, the area.[2]

The Coupling Air Chamber and Diaphragm

In order effectively to make use of horns as sound intensifiers it is usually necessary to couple the throat of the horn to the diaphragm through an air chamber. We shall first consider the effect of this air chamber on the sound output of the loud speaker. This coupling air chamber is generally of an indefinite conical shape of the type shown in Fig. 1. If we assume that this air chamber is so proportioned that

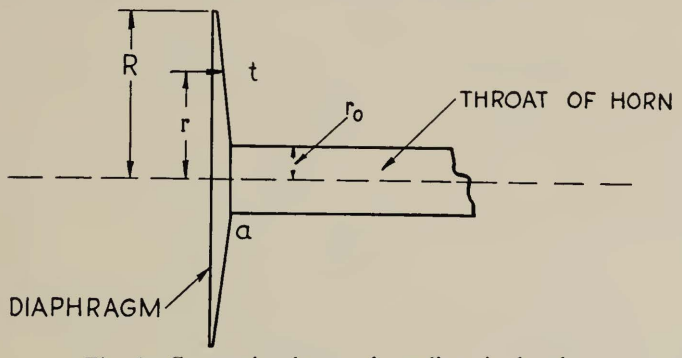


Fig. 1—Conventional type of coupling air chamber.

the annular area, $2\pi rt$, is equal to the throat area, and that the diaphragm is driven so that its displacement is paraboloidal, then, as calculated from formulæ developed in appendix A, the mechanical impedance imposed by the air chamber and horn on the diaphragm is shown in the curves of Fig. 2. Here the ordinates of the curves r_1 and x_1 are proportional to the resistance and reactance respectively, and the abscissæ are equal to the product of the frequency and the radius of the diaphragm. Of particular interest here is the large decrease in the resistance with frequency, for r_1, multiplied by the square of the velocity of the diaphragm, is the acoustic power delivered to the horn. For example, if the radius of the diaphragm were four centimeters, no sound would be emitted at 4,000 p.p.s. We have here one reason why most horn-type loud speakers fail to reproduce high frequency tones at sufficient intensity. Of course, in most cases the high frequency tones are further attenuated by the fact that the mode of motion of the diaphragm changes with frequency. The decrease in

[2] "The Function and Design of Horns," by C. R. Hanna and J. Slepian, *Journal of the A. I. E. E.*, March 1924.

resistance with frequency is largely due to the fact that the disturbances generated at different points of the diaphragm do not reach the throat of the horn in the same phase. To minimize this effect the air chamber should be designed so as to make this phase difference as small as

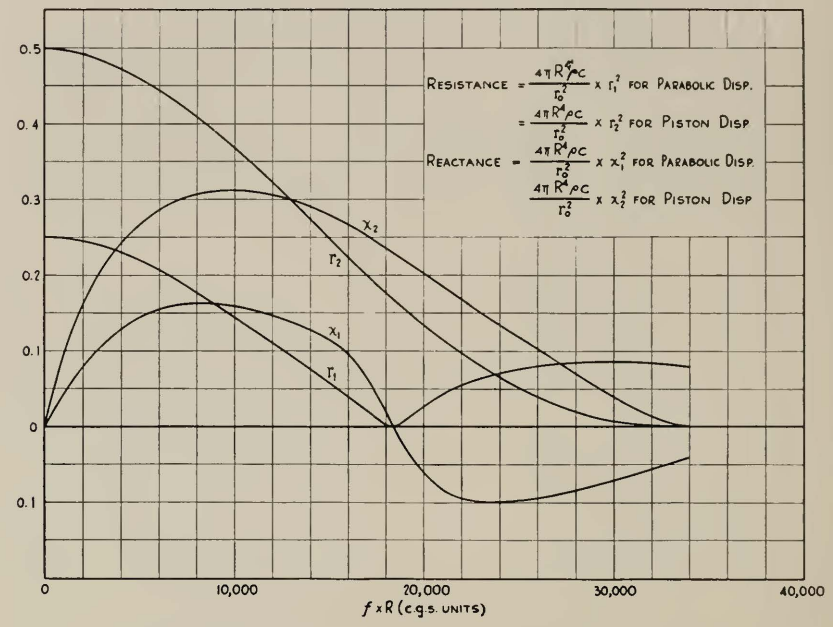


Fig. 2—Mechanical impedance of air chamber and ideal horn.

possible. In the same figure, r_2 and x_2 show the resistance and reactance respectively, if the diaphragm were moved as a plunger, i.e., with the same amplitude and phase over its whole surface. It is seen that the resistance is considerably larger and the cut-off frequency nearly twice as high. These curves show the superiority of the plunger type of diaphragm.

In order to cover the desired frequency range the method of coupling a diaphragm to the horn shown in Fig. 3 was adopted. Here the disturbances reach the horn more nearly in phase without having to pass through any restricted passages. The throat of the horn is flared annularly to the point A. The disturbances reach the throat of the horn from the inner and outer portions of the diaphragm approximately in phase up to comparatively high frequencies. With this type of construction it is possible to use a fairly large diaphragm so that large amounts of power may be delivered without a great sacrifice in efficiency at either the high or the low frequencies. An experimental test

showed that with this type of coupling for a particular size of diaphragm and throat area the cut-off frequency was raised from approximately 3,500 to 6,000 cycles per second.

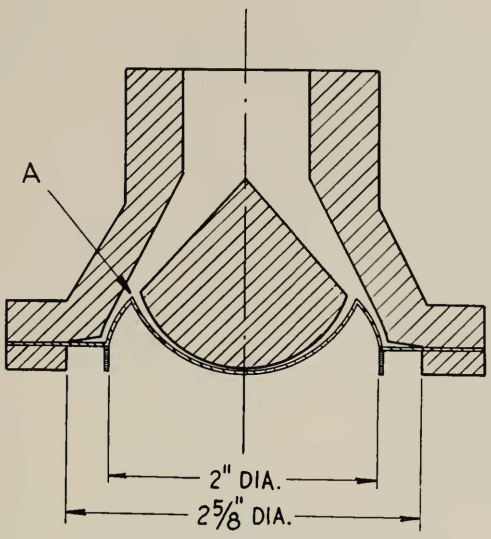


Fig. 3—Diaphragm and air chamber.

Principal Dimensions

Effective mass of coil and diaphragm = 1.0 gm.
Effective area of diaphragm = 28 sq. cm.
Area of throat of horn = 2.45 sq. cm.
Stiffness constant $= \frac{\text{force}}{\text{static displacement}} = 6 \times 10^6$ dynes/cm.
Resistance of coil = 15 ohms.
Length of wire in coil = 760 cm.
Average flux density = 20,000 gauss.

The diaphragm was made of a single piece of aluminum alloy 0.002 inch thick; metal was used in preference to other materials because of its superior mechanical properties. The form and principal dimensions are shown in Fig. 3. A driving coil is attached directly to the diaphragm near its outer edge. With this arrangement the diaphragm can be driven nearly as a plunger and it has little tendency to oscillate about a diametral axis, as there is great rigidity against a radial displacement of any part of the coil. The portion of the diaphragm lying between the coil and the clamping surfaces has tangential corrugations of the same type as described by Maxfield and Harrison [3] in reference to a phonograph sound box. The inner portion of the diaphragm was drawn into the form of two re-entrant segments

[3] *Bell System Technical Journal*, Vol. V, pp. 493–523, July 1926.

of spherical shells; this part was thereby made very rigid so that it should move as a unit up to high frequencies.

Construction of the Driving Coil

For the driving element of loud speakers either a moving coil or a moving armature is commonly used. The latter is in general satisfactory if driven at a small amplitude. However, where large powers are involved, the moving coil drive can be much more simply constructed so that it is free from amplitude distortion; it has the further advantage of having a resistance nearly constant with frequency and a practically negligible reactance. These were the primary reasons for our choosing this type of drive.

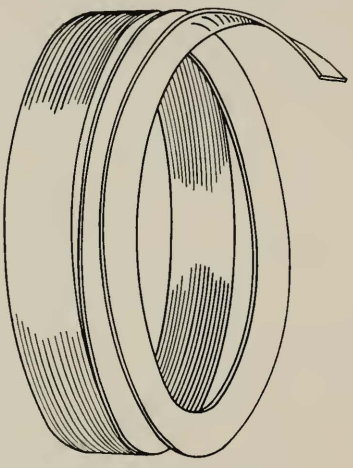


Fig. 4—Receiver driving coil.

The coil that was used in the receiver consisted of a single layer of aluminum ribbon 0.015 inch wide and 0.002 inch thick wound on edge as shown in Fig. 4. The turns were held together with a film of insulating lacquer about 0.0002 inch thick, thoroughly baked after the winding was completed. This type of coil has the following advantages. It is self-supporting, no spool being required; 90 per cent of the volume of the coil is occupied by metal; the distributed capacity between turns is small, giving a coil whose impedance varies only slightly with frequency; the metal is continuous between the cylindrical surfaces, allowing heat to be conducted rapidly outward from the center of the winding and diminishing the possibility of any warping of the coil; it can be accurately made to dimensions, thus permitting small clearances between the coil and the pole pieces. Small clearances not only permit the use of a comparatively small magnet but they facilitate the dissipation of heat. This latter effect is shown in the curves of Fig. 5. These curves give the temperature of the coil as a function of the power input for the coil in open air (*A*), and when it is placed between annular pole pieces with clearances of 0.010 inch between the cylindrical surfaces (*B*).

The Electromagnet

As shown in Fig. 6, the electromagnet is of conventional design except that the central pole piece has an opening through its center to

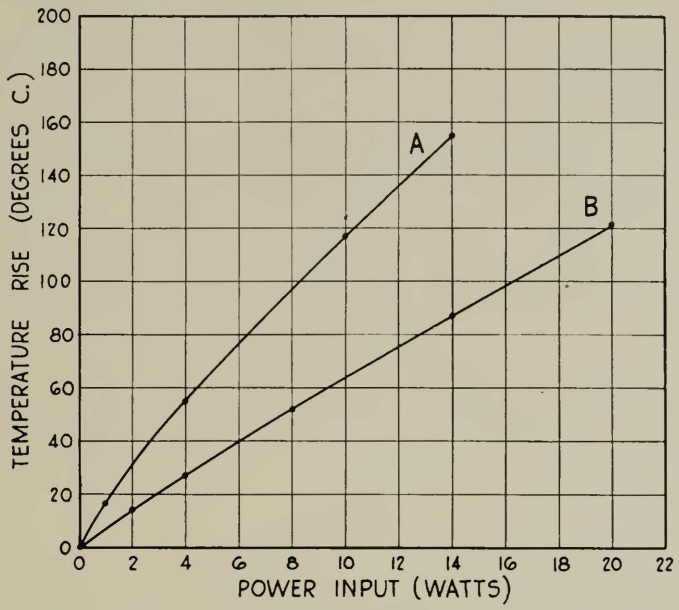


Fig. 5—Variation of temperature of coil with direct current input.

avoid a reaction of any air pockets on the diaphragm. This opening is widely flared to prevent tube resonance.

Experimental Results

It has already been pointed out that an ideal horn should have an acoustic impedance at its throat equal to that of a tube of infinite length and of the same cross-sectional area. In order to measure the

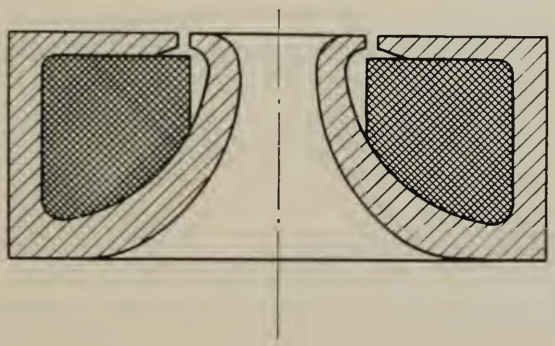


Fig. 6—Field magnet.

efficiency that the receiver would have under this condition, by any method whatever, it is necessary to connect the receiver to a tube having the same acoustic impedance as a tube of this character. This impedance for a tube 2.45 sq. cm. in area is 16.7 c.g.s. units. A tube of finite length but of the same area will have an impedance of this value provided the sound wave reflected at the far end has a relatively small amplitude when it reaches the sending end. To satisfy this condition a tube 50 feet long was terminated in an acoustic resistance unit having an impedance approximately equal to $16.7 + j16\omega \cdot 10^{-4}$ c.g.s. units. The essential elements of this resistance unit comprised a number of short narrow annular slits; its impedance was determined experimentally by a method described in another paper.[4] As this impedance at low frequencies is practically the same as the characteristic impedance of the tube, the amplitude of the reflected wave in this region is small; at the higher frequencies the reflected wave is attenuated sufficiently in the 50-foot tube to produce a negligible effect on the sending end impedance. This tube with the resistance unit was connected to the receiver during the following series of measurements.

Efficiency

One of the simplest methods of determining the power efficiency of a loud speaker is to measure the electrical impedance, first, when the receiver is in operating condition, and, secondly, when the diaphragm is constrained from moving so that no back e.m.f. is generated. The difference between these impedances is known as the motional impedance.[5] The resistance component of this motional impedance when multiplied by the square of the current gives the power that is generated by the motion of the diaphragm. If there is a negligible amount of power lost in viscosity and mechanical hysteresis, the ratio of the motional impedance to the free impedance can be taken as the efficiency of the receiver, i.e., the ratio of the acoustic power output to the total power input. This method of measuring efficiency is well known to the art, but for most commercial receivers the efficiency is so low that the motional impedance cannot be determined with a high degree of accuracy over an extended frequency range. However, for this receiver we have had no difficulty in determining the efficiency in this way up to 8,000 p.p.s. The values so obtained are given by the circles in Fig. 7.

On account of the uncertainty of the magnitude of the mechanical

[4] Wente and Bedell, *Bell System Technical Journal*, January 1928.

[5] Kennelley and Pierce, "The Impedance of Telephone Receivers as Affected by the Motion of their Diaphragms," *Proc. A. A. A. S.*, Vol. 48, No. 6, September 1912.

power losses within the receiver it was deemed desirable to measure the efficiency more directly, viz., to measure the actual sound power generated for a given power input.

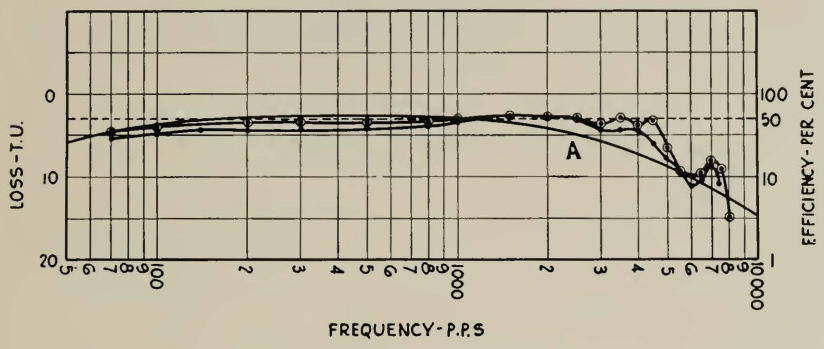


Fig. 7—Efficiency of the receiver.

The power output may be determined directly by measuring the acoustic pressure in the tube at the sending end. In order to measure this pressure an annular slit was provided on the side of the tube a few inches from the receiver as shown in Fig. 8. This annular slit had a

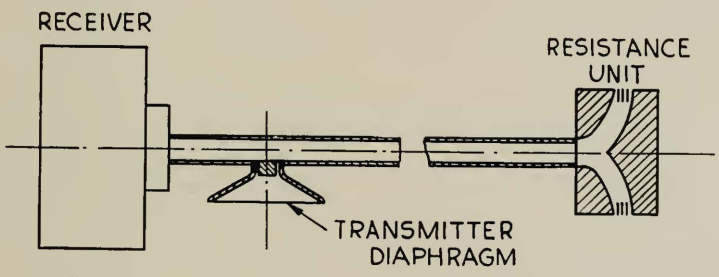


Fig. 8—Arrangement of apparatus for measuring efficiency.

diameter of a quarter of an inch, a width of 0.003 inch and a length of 0.040 inch. A slit of these dimensions has an acoustic impedance about fifteen times as great as the tube, so that it has a negligible effect on the sound wave propagated along the tube. This slit led to a small chamber over the face of the diaphragm of a condenser transmitter. The condenser transmitter was connected to an amplifier and an a.c. ammeter. This combination was previously calibrated, so that from the meter readings the pressure over the slit could be determined.

The input power was determined from the current input and the resistance of the receiver. With this set-up we thus were able to measure both the acoustic power transmitted along the tube, which

is equal to $\frac{(\text{pressure})^2}{16.7}$ ergs per second, and the power input. The operating efficiency is the ratio of these two quantities. The dots plotted in Fig. 7 give the values of the efficiencies so obtained. These values are seen to agree closely with those calculated from the motional impedance. This agreement shows that the mechanical power losses in the receiver are small.

Curve A in Fig. 7 gives the efficiency as calculated from the constants of the receiver by means of the formula given in appendix B, under the assumption that the mechanical impedance imposed on the diaphragm and the air chamber has the same value throughout the whole frequency range, viz., $16.7\,A^2$ c.g.s. units, where A is the effective area of the diaphragm. It is seen that the calculated and measured values are in good agreement except for certain irregularities at the higher frequencies. Whether these irregularities are to be ascribed to the action of the air chamber or to a change in the mode of motion of the diaphragm we are not at present prepared to say.

The curves of Fig. 7 give an efficiency for this receiver of about 50 per cent over a wide frequency range. This efficiency is within 3 T.U. of the possible maximum of 100 per cent. We may remark at this point that it is conceivably possible to build a receiver which will sound louder than one having an efficiency of 100 per cent. If, for instance, a receiver introduces harmonics on account of amplitude distortion, a low frequency driving force may give rise to a tone of higher frequency, where the ear may have a sensitivity many times greater than at the driving frequency. An increase in loudness obtained in this way of course exacts a sacrifice in the faithfulness of reproduction. The difference in loudness between the sound emitted by this receiver and by ordinary commercial types of loud speakers, for the same power input, is considerable, since most of them have an efficiency of less than one per cent for speech frequencies. Not only does this receiver have a high efficiency over a wide frequency range but it is free from any sharp variations in efficiency with frequency, a condition of great importance in the quality of reproduction.

Amplitude Distortion

Thus far we have discussed only the frequency characteristic of the receiver. There still remains to be considered the proportion of harmonics that are generated by the receiver when supplied with a current of sine wave form. These harmonics are generated when the displacement of the diaphragm is not proportional to the input current. At low frequencies the amplitude of motion for a given power output

is comparatively large and the diaphragm for large powers will be driven beyond the point where Hookes' law holds. At the higher frequencies no trouble is to be expected from this source. With the aid of an electrical filter we have therefore made measurements on the harmonic content in the sound output when the receiver was supplied with a sixty-cycle sine wave current. The values so obtained as a function of the power input are plotted in Fig. 9, where curve *A* is the

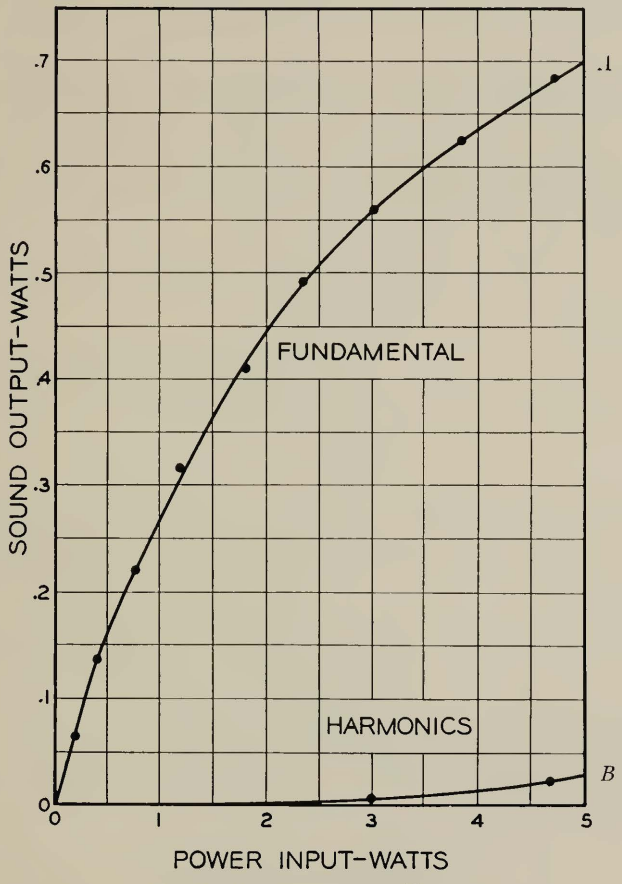


Fig. 9—Power output at 60 p.p.s.

output power of the fundamental tone and *B* that of the higher harmonics. These curves show that even at 60 cycles an output power of 0.5 watt may be obtained without the introduction of higher harmonics to an amount greater than 1.0 per cent. The total power in the harmonics would in this case be 20 T.U. below that in the fundamental

tone. If a horn were connected to the receiver in place of the tube, in addition to the resistance, a mass reactance would generally be imposed on the diaphragm at the lower frequencies. Under these conditions the proportion of harmonics introduced would be still lower than that indicated in Fig. 9.

At the higher frequencies the power output is limited solely by the current-carrying capacity of the coil. At these frequencies the steady power input for a temperature rise of 100 degrees C. is about 30 watts. With an efficiency of 50 per cent the corresponding output would be 15 watts.

After the work described in this paper was for the most part done and as a result of the extremely promising performance of the first models, a design of the receiver built along essentially these lines was worked into a form suitable for commercial production by Mr. W. C. Jones and Mr. L. W. Giles. These receivers are now in commercial use in Vitaphone and Movietone installations. As commercially produced in quantities numbering several thousand, efficiencies of the order of 30 per cent have been realized.

In conclusion, we wish to express our appreciation for the valuable assistance given by Mr. T. F. Osmer in carrying out most of the experimental work described in this paper.

Appendix A

Consider a diaphragm and connecting air chamber of the form shown in Fig. 1. Assume that the air chamber is of a form such that the cross-sectional area at any distance r from the center is equal to the throat area of the horn, i.e., $2\pi rt = \pi r_0^2$. This form of connecting air chamber then differs but little from that used in most commercial types of horn speakers. The sound output is in general dependent on the mode of motion of the diaphragm. In most loud speakers this mode of motion varies with the frequency. However, let us assume that we have a paraboloidal displacement at all frequencies. The velocity at any radial distance may then be represented by

$$\xi = \xi_0 \left[1 - \left(\frac{r}{R} \right)^2 \right] e^{i\omega t}$$

if $\xi_0 e^{i\omega t}$ is the velocity at the center.

Under the assumed conditions, the sound transmitted through the throat is very nearly the same as that which would be transmitted along the positive direction through the tube sketched in Fig. 10, which extends to infinity in both directions, provided the portion of the wall

of the tube from a' to 0 and from a to 0 had a radial velocity equal to

$$\frac{2\pi r}{2\pi r_0} \cdot \dot{\xi}_0 \left[1 - \frac{r^2}{R^2} \right] e^{i\omega t}.$$

The velocity potential at a point, P, at a distance y from a, if r_0 is small compared with the wave-length of sound, is then

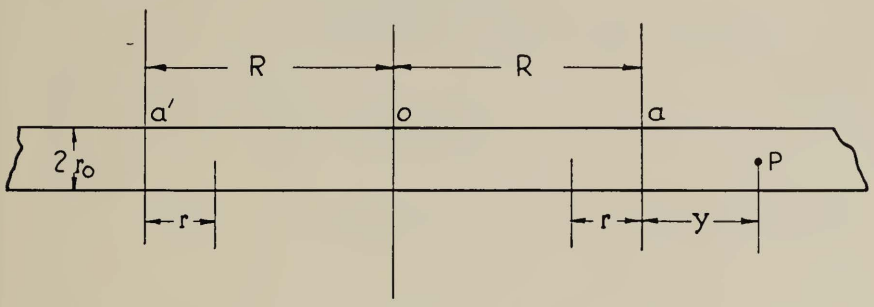


Fig. 10.

$$\varphi_y = -i \left[\int_0^R \frac{\left[1 - \frac{r^2}{R^2} \right] r}{r_0^2 k} e^{ik(ct-y-r)} dr + \int_0^R \frac{\left[1 - \frac{r^2}{R^2} \right] r}{r_0^2 k} e^{ik(ct-y-2R+r)} dr \right] \dot{\xi}_0,$$

where c is the velocity of sound and

$$k = \frac{\omega}{c}$$

or

$$\varphi_y = A \frac{2\dot{\xi}_0 e^{ik(ct-y-R)}}{-ir_0^2 k^3},$$

where

$$A \equiv \left(1 + \frac{6}{k^2 R^2} \right) \cos kR + \left(2 - \frac{6}{k^2 R^2} \right).$$

If we take the real part of

$$\rho \frac{d\varphi_y}{dt},$$

we get the instantaneous pressure at the point, P, where ρ is the density of the air. This gives

$$p_y = \frac{-2\dot{\xi}_0 \rho c}{r_0^2 k^2} \cdot A \cos k(ct - y - R).$$

If we substitute r for $-y$, this expression gives the instantaneous pressure on the diaphragm at the radial distance r from the center. The instantaneous power delivered by the diaphragm is then

$$W = -\frac{2\rho cA}{r_0^2k^2}\dot{\xi}_0 \cos kct \cdot 2\pi \int_0^R \left(1 - \frac{r^2}{R^2}\right) \cos k(ct + r - R) \cdot r dr$$

$$= \frac{4\pi A\rho c}{r_0^2k^4}\dot{\xi}_0^2[A \cos kct + B \sin kct] \cos kct,$$

where

$$B \equiv \left(1 + \frac{6}{k^2R^2}\right) \sin kR - \frac{6}{kR}.$$

The effective force on the diaphragm is then

$$F = \frac{W}{\dot{\xi}_0 \cos kct} = \frac{4\pi A\rho c\dot{\xi}_0}{r_0^2k^4}[A \cos kct + B \sin kct].$$

The effective resistance is therefore

$$\frac{4\pi A^2\rho c}{r_0^2k^4}$$

and the effective reactance

$$-\frac{4\pi\rho cAB}{r_0^2k^4}.$$

Expressions for the resistance and reactance for other modes of motion of the diaphragm may be obtained in a similar manner.

Appendix B

The motional resistance, R_m, of a moving coil receiver is equal to

$$\frac{B^2l^2(r + r')}{(r + r')^2 + \left(m\omega - \frac{S}{\omega} + x\right)^2}$$ ohms,[6]

where B is the average flux density,

l, the length of wire in the receiving coil,

$r + jx$, the mechanical impedance imposed on the diaphragm by the horn through the coupling air chamber,

[6] Kennelley and Pierce, loc. cit.

$r' + j\left(m\omega - \frac{S}{\omega}\right)$, the mechanical impedance of the diaphragm aside from that imposed by the horn.

If r' is negligible, the efficiency of the receiver, i.e., the ratio of power output to power input, is

$$\eta = \frac{R_m}{R_m + R_d},$$

where R_d is the resistance of the coil when its motion in the field is completely damped.

Abstracts of Bell System Technical Papers Not Appearing in this Journal

A Photo-electric Process of Halftone Negative Making Applicable over Telephone Lines.[1] H. E. IVES. The commercial process of making halftone engravings breaks the picture up into a large number of dots by means of a screen. Unless special intermediate processes are adopted this does not reproduce the tones correctly because the size of the dots is not directly proportional to the light intensity. This paper describes an adaptation of the photo-electric system of picture transmission, as an alternative for use of screen, which produces individual dots more accurately proportional in size to the light intensity.

The method proposed thus affords a means of improving quality in halftone engraving by using an outfit similar to the picture transmitting and receiving outfit. Used in connection with the commercial picture transmission service, it can provide pictures with an accurate tone structure during the transmission process so that the resulting copy is ready for the engraver without the use of any screen process. Full details of several arrangements of the apparatus are given together with engravings produced by the photo-electric method.

The picture transmission system as used transmits the picture in the form of a continuous strip of varying intensity. By the introduction of a synchronized sectored disc this strip is made discontinuous, forming the dots for the halftone process.

Advance Planning of the Telephone Toll Plant.[2] J. N. CHAMBERLAIN. A general review of the commercial studies which precede the design of telephone toll plant is given together with some specific data concerning the toll line conditions in the northern California area of the Pacific Telephone and Telegraph Company. Attention is called to the special conditions requiring a large amount of submarine cable plant resulting from the peninsular location of San Francisco. The Pacific Company has plans for about 1,000 miles of toll cable network in its present program which it expects to install at the rate of 100 miles a year. The article places special emphasis on the commercial factors governing the choice between cable and open wire construction for future extensions of the toll plant.

[1] *Opt. Soc. Amer. Jl.*, Vol. 15, p. 96, August, 1927.

[2] *Jl. Am. Inst. El. Engrs.*, Vol. 46, p. 994, October, 1927.

The Adsorption of Gases by Solids with Special Reference to the Adsorption of Carbon Dioxide.[3] H. H. LOWRY and P. S. OLMSTEAD. This paper presents a theory of adsorption and its mathematical development, which is similar to but differs somewhat from that of Polanyi. A detailed description is included of a relatively convenient method of application of the theory to experimental data. Using this procedure, a test of the theory has been made on the data obtained by Homfray, Titoff, Richardson, Chappuis and S. O. Morgan for the adsorption of carbon dioxide by charcoal. The very satisfactory agreement obtained between experiment and theory gives support to the fundamental assumptions underlying the theory.

The Densities of Coexisting Liquid and Gaseous Carbon Dioxide and the Solubility of Water in Liquid Carbon Dioxide.[4] H. H. LOWRY and W. R. ERICKSON. The densities of coexistent liquid and gaseous carbon dioxide are measured over the temperature range − 5.8 to 22.9° and it is shown that they can be satisfactorily represented by equations involving the first and one third powers of the temperature on the critical scale. The data are shown to be in substantial agreement with those of other observers. It is also shown that the density of saturated carbon dioxide vapor is the same within the experimental error in the presence or absence of water, from which it is concluded that the solubility of water in liquid carbon dioxide is less than about 0.005 per cent by weight over the temperature range of the investigation. Attention is called to qualitative evidence of the formation of a solid hydrate of carbon dioxide at about 4°.

Atomic Grouping in Permalloy.[5] L. W. MCKEEHAN. This is a theoretical paper which follows a long series of experimental papers. In a solid solution of two metals, e.g., in the solid solution of nickel and iron known as permalloy, the atoms of both kinds occupy in each crystal the points of a single space-lattice. It is important to know whether the points so occupied by atoms of a single kind are located at random or have some regularity of arrangement. If the latter is the case, it may be asked further whether the regularity is due to the frequent occurrence of definite groupings of unlike atoms or merely to a tendency for atoms of one kind to separate from each other as widely as possible. The magnetic properties of permalloy are here taken to show that the last-mentioned possibility is probably nearest

[3] *Journal of Physical Chemistry*, Vol. 31, pp. 1601–1626, November, 1927.
[4] *Journal of the American Chemical Society*, Vol. 49, pp. 2729–2734, November, 1927.
[5] *Jl. Franklin Inst.*, 204, 501–524 (1927).

to the truth. The method is a combination of analytical and graphical analysis applied to several hypothetical solid solution crystals each containing more than a thousand atoms. It may also, as is pointed out in the paper, be used in problems concerning other than magnetic properties of solid solutions.

The Short Wave Limit of Vacuum Tube Oscillators.[6] C. R. ENGLUND. A study of the shortest attainable undamped waves which can be produced with vacuum tube oscillators resulted in the production of one-meter waves as the fundamental mode of oscillation of the circuit. The shortest waves "attainable with reasonable ease" with several types of standard tube are found to be:

Tube	Meters
W. E. 230-D (60 mil. fil.)	2.0
" 205-D ("E" 5-watt)	3.2
" 221-D (1/4 amp. fil.)	3.3
" 211-D ("G" 5-watt)	3.5

In order to reduce the capacity as far as possible some experiments were made with unbased tubes. A special 5-watt tube produced four-meter signals which were received up to one-mile distances. An interesting feature of the work was the interference caused by the presence of the observer while conducting the experiments, because of the action of the human body as a tuned antenna at these wavelengths.

A General Theory of the Correlation of Time Series of Statistics.[7] M. K. ZINN. In mathematical physics elaborate methods have been devised for dealing with oscillatory systems, damped or undamped, like pendulums, vibrating strings, vibrating telephone diaphragms, and with systems that are essentially damped but have no oscillatory characteristics, like the flow of heat and diffusion. The writer of this article approaches the problem of the economic structure with a point of view which sees the business world as behaving like such an oscillatory system.

The structure of economic society is a system exhibiting certain structural factors which express the tensions that are set up by a shift in prices here on some other factor like employment there. The problem is: How are these tensions to be inferred from statistics which describe the observed behavior of the economic world? The economist, unfortunately, cannot disconnect, say, the Federal Reserve system from the rest of the economic structure, connect it to a portable test

[6] *Proc. I. R. E.*, 15, 914, November, 1927.
[7] *The Review of Economic Statistics*, Harvard Economic Service, *9*, 184 (1927).

set and read off its economic impedance in the same way that an engineer can test a transformer.

He has to take instead the observed fluctuations in time of two series, say "wholesale commodity prices" and "commercial paper rates," and try to infer from such data the way in which changes in one of these quantities react on the other. The paper is concerned with the general way of doing this with a full analysis of the relation between these two economic variables from this point of view.

Just as most of the theory of oscillatory systems in mathematical physics is confined to linear systems, so the writer finds it convenient to assume linear relations between the economic variables. This greatly simplifies the mathematics. The formulas come out to be analogous to some formulas of electric circuit theory when approached from the Heaviside operational standpoint. It thus appears that much alternating current theory may come to be of value in studying economic variations. It is pleasant to think that some years hence we may be using the language of "a.c." theory (reluctance, susceptance, impedance, etc.) to describe the functional relation of one economic unit to the rest of society. Perhaps to study the relation of conditions in one part of the country to those in another we shall be using the long line transmission theory. Perhaps the theory will show us how to build economic band-pass filters which will protect us from too great fluctuations in business conditions, etc.

The application of the ideas of oscillatory systems to economics, here well started, is a subject which, it is believed, will strike a responsive chord in the heart of every electrical engineer and every mathematical physicist.

Contributions of Chemical Science to the Communications Industry.[8] CLARENCE G. STOLL. The author considers the improvements that have been made under a four-fold grouping of materials into electrically conducting, magnetic, insulating, and materials for apparatus structures. Particular emphasis is laid on the influence the chemist has exerted on the control of materials for manufacturing purposes. So many undesirable variations in manufactured products are caused by lack of uniformity in the raw materials that the aid of the chemist in standardizing testing and sampling methods cannot well be overestimated.

In conclusion, he pays homage to the readiness with which chemists have responded to demands for improvement in many of the raw materials and suggests that because of it similar demands may be of

[8] *Journ. Ind. and Engr. Chem.*, Vol. 19, p. 1132, 1927.

more frequent occurrence in the future. Among such possible innovations he mentions a better conductor of electricity, a cable covering superior to the present lead alloys, higher grade insulation, and a contact material less subject to corroding and freezing. The recent advances made in metallic alloys give every promise of a more brilliant future.

The Thickness of Spontaneously Deposited Photoelectrically Active Rubidium Films, Measured Optically.[9] H. E. Ives and A. L. Johnsrud. Measurements of the phase shift on reflection of polarized light from a reflecting surface on which there is deposited a very thin film of metal are described. The materials were the thin films of rubidium which are slowly deposited on glass or platinum when these have been thoroughly out-gassed. The apparatus was so arranged that measurements of the photo-electric activity could be made simultaneously with determinations of the optical effect due to the thin film. The data were then interpreted in terms of the electromagnetic theory of light using special developments due to T. C. Fry, to be published in the *Journal of the Optical Society* for January, 1928.

It is concluded from the measurements that the film of rubidium deposited on glass after fourteen days is of the order of magnitude of one atom thick. The theoretical effect of a layer of rubidium on platinum of the order of even several atoms thick is, however, so small as to lie within the errors of observation. "It thus appears, if the validity of the optical measurements of thickness is conceded, that the photo-electric emission is obtained when a layer of rubidium of approximately one atom in thickness is present," they conclude.

[9] *Opt. Soc. Amer. Jl.*, Vol. 15, 374, December, 1927.

Contributors to this Issue

E. C. Wente, A.B., University of Michigan, 1911; S.B. in Electrical Engineering, Mass. Inst. of Technology, 1914; Ph.D., Yale University, 1918; instructor in physics and mathematics, Lake Forest College, 1911–12; Engineering Department, Western Electric Company, 1914–16, 1918–24; Bell Telephone Laboratories, 1924–. Mr. Wente has worked principally on general acoustic problems and on the development of special types of acoustic devices.

E. H. Bedell, S.B., Drury College, 1924; University of Missouri, 1924–25; Bell Telephone Laboratories, Inc., 1925–. Since coming to the Laboratories Mr. Bedell has studied various acoustic problems, notably those of an architectural character.

John R. Carson, B.S., Princeton, 1907; E.E., 1909; M.S., 1912; Research Department, Westinghouse Electric and Manufacturing Company, 1910–12; instructor of physics and electrical engineering, Princeton, 1912–14; American Telephone and Telegraph Company, Engineering Department, 1914–15; Patent Department, 1916–17; Engineering Department, 1918; Department of Development and Research, 1919–. Mr. Carson is well known through his theoretical transmission studies and has published extensively on electric circuit theory and electric wave propagation.

Paul P. Coggins, Harvard University, A.B. in Mathematics, 1920, A.M. in Physics, 1921; Department of Development and Research, American Telephone and Telegraph Company, 1921–27. Statistician, New Jersey Bell Telephone Company, October 1927–. Up to October 1, 1927, Mr. Coggins dealt with the application of the mathematical theory of probabilities, including sampling theory, to various telephone problems.

W. J. Shackelton, B.S. in E.E., University of Michigan, 1909; Western Electric Company, Manufacturing and Installation Department, 1909–10; Bell Telephone Laboratories, 1910–. Mr. Shackelton's principal activities have been in connection with the design of loading coils and the development of methods of high frequency measurement.

J. G. Ferguson, B.S., University of California, 1915; M.S., 1916; research assistant in physics, 1915–16; Bell Telephone Laboratories,

1917–. Mr. Ferguson's work has been in connection with the development of methods of electrical measurement.

C. J. Davisson, B.Sc., University of Chicago, 1908; Ph.D., Princeton University, 1911; instructor in physics, Carnegie Institute of Technology, 1911–17; research engineer, Western Electric Company and Bell Telephone Laboratories, 1917 to date. Dr. Davisson's work since coming with the Bell System has related largely to thermionics and electronic physics.

E. Peterson, Cornell University, 1911–14; Brooklyn Polytechnic, E.E., 1917; Columbia, A.M., 1923, Ph.D., 1926; Electrical Testing Laboratories, 1915–17; Signal Corps, U. S. Army, 1917–19; Engineering Dept., Bell Telephone Laboratories, 1919–.

C. R. Keith, B.S., 1922, California Institute of Technology; M.A., 1925, Columbia University; Carrier Research Department, Bell Telephone Laboratories, 1922–. Mr. Keith's work has related to the study of vacuum tube and magnetic modulators and other carrier apparatus.

A. L. Thuras, B.S., University of Minnesota, 1912; E.E., 1913; laboratory assistant with U. S. Bureau of Standards, 1913–16; graduate student in physics, Harvard, 1916–17; scientific observer with U. S. Coast Guard, 1917–19; oceanographer, 1919–20; Bell Telephone Laboratories, 1920–. Since joining the Laboratories staff, Mr. Thuras' work has had to do largely with mechanical impedance studies and bridges.

The Bell System Technical Journal

April, 1928

Joint Meeting of the Institution of Electrical Engineers and the American Institute of Electrical Engineers

ON February 16th last, a joint session of the Institution of Electrical Engineers in London and the American Institute of Electrical Engineers in New York was made possible by the transatlantic telephone. The audience in New York numbered over one thousand and that in London was several hundred. The two audiences were called to order at 10:30 A.M. New York time by Mr. Bancroft Gherardi, president of the American Institute of Electrical Engineers, and several papers were read both in New York and London to which the two audiences listened simultaneously. This joint meeting marks such an important milestone in the history of electrical communication that its entire proceedings are reproduced herewith. They are entirely self-explanatory.

Having called the meeting to order, Mr. Gherardi said:

"I will ask Mr. Charlesworth, Chairman of our Meetings and Papers Committee, to say a few words concerning the London meeting, and then to arrange for our joint session."

MR. CHARLESWORTH: Before proceeding with the joint session with our associates in the British Institution of Electrical Engineers, I wish to say just a few words concerning their London meeting in order to help you visualize the nature and significance of our joint session.

Our British associates have assembled in the auditorium of the Institution of Electrical Engineers Building located on the Victoria Embankment. The time is about 3:30 in the afternoon. Their meeting includes their President, Archibald Page, Chief Engineer of Central Electricity Board, their Vice President, Colonel Purves, Engineer and Chief of the British Post Office, the full Council of the Institution, members and invited guests from all parts of Great Britain, men prominent in all branches of the electrical industry.

Through the medium of developments which have been made in electrical communication, we are in effect to wipe out the great distance which separates the meeting places of our two societies, and to come together in a joint session in which our respective Presidents may exchange greetings in our behalf and in which other distinguished

representatives of our two societies may take part. By means of the loud speakers we shall be able to hear these proceedings as though we were all located in one great auditorium.

Colonel Purves has just finished making a statement to his associates concerning our meeting here in New York.

I will now speak to Colonel Lee who is at the telephone in London, and we will then proceed with our joint session.

Good morning, Colonel Lee.

COLONEL LEE: "Good afternoon, Mr. Charlesworth."

MR. CHARLESWORTH: "Are we ready to proceed with our joint session, Colonel Lee?"

COLONEL LEE: "We are, Mr. Charlesworth."

MR. CHARLESWORTH: "I will hand the telephone to Mr. Bancroft Gherardi, President of the American Institute of Electrical Engineers."

COLONEL LEE: "I will also hand the telephone to Mr. Archibald Page, President of the British Institution of Electrical Engineers."

MR. GHERARDI: Good morning, Mr. Page.

MR. PAGE: Good afternoon, Mr. Gherardi.

MR. GHERARDI: Mr. Page, it would give us great pleasure, if as President of the Institution of Electrical Engineers—the senior society, founded in 1871—you would act as chairman of this joint meeting.

CHAIRMAN PAGE: I regard it as a great honour to be asked to take the chair on this historic occasion. It is also a gracious compliment to our institution, and in accepting, which I do gladly, I desire to thank you, Mr. President, and the members of the American Institute of Electrical Engineers most heartily. I welcome all present at the meeting now in session, and venture to predict that the proceedings will prove exceedingly interesting and likely to live not only in our memories, but to be quoted by succeeding generations of electrical engineers as marking an important milestone in the advancement of electrical science. I am sure I interpret the desire of those assembled if I request Mr. Gherardi to address us, which I now do.

MR. GHERARDI: Mr. President and Members of the Institution of Electrical Engineers: On behalf of the American Institute of Electrical Engineers, I extend to you greetings and our best wishes. We are meeting here in New York at our Midwinter Convention. In the auditorium of the Engineering Societies Building in New York City, from which I am speaking, there are assembled about one thousand members of our organization from all parts of the United States, from Canada, and from other parts of the New World. It is with the greatest satisfaction that, as a result of the accumulated work of the scientist, the inventor, and the electrical engineer, it is possible for us

to hold this joint meeting—the first of its kind. It is with feelings of deep appreciation and respect that we think of the men who have exemplified the ideals of your organization—Faraday, Maxwell, Kelvin—and of the many others, past and present, who have contributed to Electrical Engineering and to the scientific foundations upon which it rests. These developments have been notable and have contributed in the greatest degree to the welfare of mankind. One of these developments is the art of electrical communication—the electric telegraph, and the telephone. These have made communication independent of transportation and no longer subject to all of its difficulties and delays. By the telephone, distance has not only been annihilated, but communication by means of the spoken word has become possible. Starting in 1876 with instruments and lines which, with difficulty, permitted communication over distances limited to a few miles, the telephone art has been improved year by year until continents have been spanned and, at last, even the limitations of the Atlantic Ocean have been overcome, and today telephone conversation between the two great capitals of the English-speaking world is a reality. We are gratified that transatlantic communication has made this meeting possible; it has added one more to the many ties existing between our two institutions and has added still another opportunity for friendly communication between us.

CHAIRMAN PAGE: Mr. Gherardi and gentlemen: Please regard me for the time being, not as chairman but rather as representing the thirteen thousand members of the Institution of Electrical Engineers. My first desire is to thank you, sir, for your most kind message of good will to us all. In turn we hail the President and members of the American Institute of Electrical Engineers with feelings of the utmost warmth and of everything included in the term good comradeship. The telephone must rank as one of the greatest inventions of the nineteenth century and it has transformed the daily life of all civilized people. Our indebtedness to Graham Bell for the boon he has conferred upon us increases with the years, and his memory, along with that of Franklin and Henry, will be cherished as becomes such benefactors of mankind. It would indeed be a gigantic task to attempt to exhaust the list of those of your society who have contributed so largely to the progress of electrical science and I must content myself by paying tribute to a great institution which has given proof time and again that engineering is truly international. It cannot be questioned that we are living in a period of extraordinary change due to scientific discovery, and in no field has the advance been more marked than in that of communication engineering. The commercial radio services

thus placed at our disposal assure closer and closer association between the English-speaking races, new spice is added to life and bonds of friendship materially strengthened. I rejoice that our two institutions can combine in the future even more effectively than in the past and that this is the outcome of the splendid work done in one of the branches of our own profession. I will now resume my chairmanship and call upon Dr. Jewett to speak.

Dr. Jewett: Mr. Chairman, Mr. Gherardi, and fellow members of the Institution of Electrical Engineers and of the American Institute of Electrical Engineers: The opportunity which this occasion offers of addressing jointly two widely separated groups of engineers whom, in times past, I have addressed vis-a-vis in London and New York, affords me the liveliest satisfaction.

I am gratified to participate in an event which marks both a notable advance in electrical communication and a pioneer demonstration of a wider use for electrical communication.

I am frankly pleased that, in common with numerous associates on both sides of the Atlantic, it has been my good fortune to play a part in the development work which has made this occasion possible.

Col. Purves and Mr. Gherardi will remember, and the rest of you will be interested to know, that in London more than a year ago, when we were engaged in final considerations preliminary to the opening of commercial transatlantic telephony, we discussed the details of just such a meeting as this. That our discussion should have been serious and not a pleasant mental diversion at a time when the channels of communications were not in operation is a striking evidence of the sound basis which underlies present-day electrical engineering. The fact that we saw and appraised the many obstacles to be overcome did not in the least diminish the assurance with which we talked of and planned for a distant event.

While therefore the present occasion is highly gratifying to the engineers whose work has made it possible, it is in no sense a surprise.

The success of this occasion is significant also in that it is the tangible evidence of a cooperation both intimate and full between men so situated as to make cooperation difficult. On behalf of my associates in America, I salute our associates in England.

Chairman Page: It is now Colonel Purves' turn to speak.

Colonel Purves: Mr. President, Mr. Gherardi, Dr. Jewett and gentlemen: It is an honour and a privilege to be associated with this notable event, which, one can justly feel, is breaking new ground in the advance of nations towards closer relationship. It is a great thing that two large assemblies, separated by wide expanses of ocean, can

join themselves together as we are doing now and interchange their thoughts and ideas by the simple and natural medium of direct speech to a combined audience. It opens up the prospects of results which thrill the imagination, and which are bound to be beneficent, and to conduce, by the way of clearer and mutual understanding, to the good of mankind. On this first occasion it is inevitable that the many professional interests which our two institutions share and which we should dearly like to talk over with each other should be pushed a little into the background, and that we should find ourselves preoccupied mainly with the wonder of the thing itself.

The radio art has given us its essential basic principles and the high power amplifying tubes, which over here we call valves. Long distance telephony has contributed a host of new devices which are equally essential. Specialized broadcast has given us the loud speaking receiver. As we sit and talk to each other our speech is launched into the air by the radio transmitting stations at Rugby and at Rocky Point with an electromagnetic wave energy of more than two hundred horsepower, and, I may add, the combined effect of the various refinements and special devices included in the transmitting and receiving systems is to make the speech efficiency of each unit of this power many thousands of times greater than that of an equivalent amount of power radiated by an ordinary broadcasting station. Many further improvements are being studied.

I should like to express the feelings of great personal pleasure with which I am listening to the voices of my old and valued friends of the American Telephone and Telegraph Company, Mr. Gherardi, Dr. Jewett and General Carty, and to assure them and their colleagues, both on my own behalf and on behalf of the engineering staff of the British Post Office, that the increased opportunities of cooperation with them which the development of the transatlantic telephone system has afforded us, are appreciated in a very high degree. We have to thank them for much helpful counsel in this and in many other matters and we look forward with great pleasure to a continuance of our close association with them on the long road forward, over which we still have to travel together.

CHAIRMAN PAGE: We are delighted to have with us in New York General John J. Carty, Vice President of the American Telephone and Telegraph Company and Past President of the American Institute of Electrical Engineers. It gives me great pleasure indeed to have this opportunity to congratulate General Carty on the presentation which he received last evening of the John Fritz Medal. This was presented to him by the National Engineers Societies of the United States for

his outstanding achievements in the engineering field. General Carty is widely regarded as the *doyen* or, to be more correct, the dean of the telephone engineering profession, and we shall be glad if he will say a few words and propose a resolution on the subject of our joint meeting.

General Carty spoke for a moment and then offered the following resolution.

WHEREAS on this 16th day of February, 1928, the members of the Institution of Electrical Engineers assembled in London, and the members of the American Institute of Electrical Engineers assembled in New York, have held, through the instrumentality of the transatlantic telephone, a joint meeting at which those in attendance in both cities were able to participate in the proceedings and hear all that was said, although the two gatherings were separated by the Atlantic Ocean; and as this meeting, the first of its kind, has been rendered possible by engineering developments in the application of electricity to communication by telephone; therefore,

Be it resolved that this meeting wishes to express its feelings of deep satisfaction that, by the electrical transmission of the spoken word, these two national societies have been brought together in this new form of international assembly, which should prove to be a powerful agency in the increase of good will and understanding among the nations; and

Be it further resolved that a record of this epoch-making event be inscribed in the minutes of each society.

CHAIRMAN PAGE: Sir Oliver Lodge, who needs no introduction, is sitting beside me and I have asked him to second the motion.

SIR OLIVER LODGE: Mr. Chairman, I think it very kind of you and the Council to allow me to take part in this important occasion, to send greetings to our many American friends. It is surely right and fitting that a record of the transmission of human speech across the Atlantic be placed upon the minutes of those societies whose members have been most instrumental in making such an achievement possible, and I second the proposal that has just been made from America. All those who in any degree have contributed to such a result from Maxwell and Hertz downwards, including all past members of the old British society of telegraph engineers, will rejoice at this further development of the power of long distance communication. Many causes have contributed to make it possible; that speech is transmissible at all is due to the invention of the telephone. That speech can be transmitted by ether waves is due to the invention of the valve and the harnessing of electrons for that purpose. That ether waves are constrained by the atmosphere to follow the curvature of the earth's surface is an unexpected bonus on the part of Providence, such as is sometimes vouchsafed in furtherance of human effort.

The actual achievement of today, at which we rejoice and which posterity will utilize, must be credited to the enthusiastic cooperation owing to the scientific and engineering skill of many workers in the background whose names are not familiar to the public as well as to those who are well known. The union and permanent friendliness of all branches of the English-speaking race, now let us hope more firmly established than ever, is an asset of incalculable value to the whole of humanity. Let no words of hostility be ever spoken.

CHAIRMAN PAGE: Gentlemen, you have heard the motion proposed by General Carty and seconded by Sir Oliver Lodge. I now put it to the joint meeting. Those in favor. Contrary. Carried unanimously. I suggest, Mr. Gherardi, that we now adjourn the meeting. I feel that it has been eminently successful and that we should regard it as the forerunner of many more to come.

PRESIDENT GHERARDI: Mr. Page, before we adjourn, I should like to take this opportunity to thank you for the gracious manner in which you have acted as chairman of this meeting, the first of its kind that ever has been held. We on this side send you our goodbye greetings and consent to the adjournment of the meeting.

CHAIRMAN PAGE: That is all the business, gentlemen. The meeting is adjourned. Goodbye.

Transatlantic Telephony—the Technical Problem

By O. B. BLACKWELL

SYNOPSIS: This paper, which, as it was read, was prefatory to the joint meeting, describes in rather non-technical terms the engineering problems involved in developing the transatlantic radio trunk by means of which the American telephone system of some 18,000,000 stations can communicate with the English telephone system of about 1,500,000 telephones, and also with the telephone systems of other European countries.

WE wish to give you a picture, necessarily very briefly sketched, of the physical makeup of the transoceanic telephone circuit, why it has been given its present form, and what further improvements are expected as the result of development work now under way.

The problem in brief is suggested by Fig. 1. A telephone system in America of some 18,000,000 stations, and distances of upwards of 3,000



Fig. 1—Map showing U. S. and European telephone systems separated by ocean

miles. A telephone system in England of about 1,500,000 telephones and the possibilities already partly realized of wire extensions to the other European nations. Three thousand miles of ocean between these two systems.

The establishment of a connection across the ocean presented two problems. First, the problem of setting up the radio circuit between the United States and England and second, the problem of making this radio circuit function as a link between these two widely extended telephone systems.

Fig. 2 shows the geographical layout of the long wave transoceanic circuit. The course followed by the currents in a connection is as follows: voice currents originating at any substation in America are

transmitted to New York City over the wire circuits in the usual way and thence also by wire to the sending station at Rocky Point, Long Island, where they are radiated into space. These waves are picked

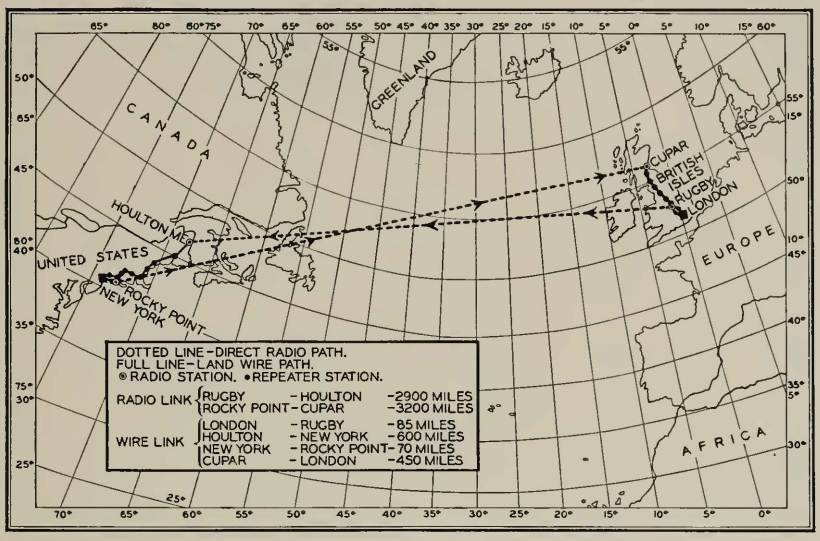


Fig. 2—Showing long wave routes

up at Cupar, Scotland, and transmitted by wires to London from which point they go by the usual wire connections to the subscriber in England or on the continent.

The answering voice waves are transmitted from the European subscriber by wire to London and thence by wire to Rugby, England, at which point they are radiated into space. The waves are picked up at Houlton in the northern part of Maine and transmitted by wires to New York City and thence by wires to the American subscriber.

You will note that the east and westbound radio systems are entirely separate from each other. Please note also that the receiving points in both countries are carried as far north as convenient—to Houlton, Maine, in this country, and to Cupar, Scotland, in the British Isles.

The radio and wire plant in Great Britain is owned and operated by the Post Office Department of the British Government.

As a supplement to the long waves there is a short wave circuit being formed which is so far only partially in use. In Fig. 3 the heavy lines show the long wave and the lighter lines the short wave routing. The short wave circuit from the United States to London was employed as an emergency routing during the severe static season last

summer extending from Deal Beach, N. J., to New Southgate, London. The British Post Office in January started sending on a return short

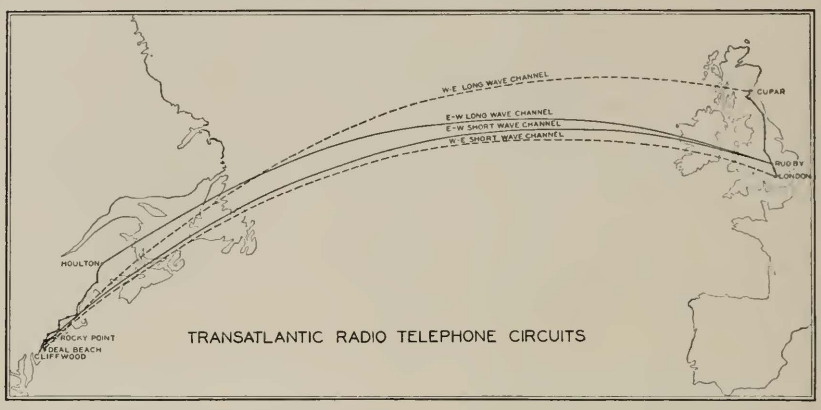


Fig. 3—Showing both long and short wave routes

wave circuit from Rugby, England, which is now being received at Cliffwood, N. J., but is not yet ready for service.

The right-hand drawing of Fig. 4 shows, necessarily on a logarithmic scale, approximately the frequency ranges covered by radio as we now know it. At the lower end are the long waves used in long distance

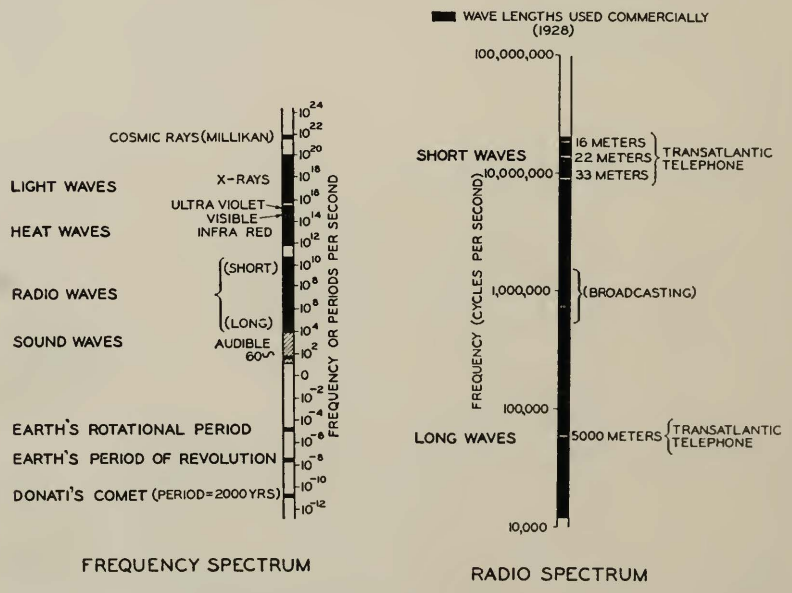


Fig. 4—Showing two plots on a logarithmic scale, one the radio frequency range, the other a general plot of frequencies

telegraphy extending down to nearly 10,000 cycles. At the upper end are the short waves already more or less exploited extending to about 10 meters, that is, 30,000,000 cycles.

It is interesting to note that one frequency range around 60,000 lying near the lower end of the scale and another frequency range extending from about 10,000,000 to 20,000,000 near the upper end of the present scale appear to be the most suitable for transoceanic transmission.

The left-hand figure has no bearing on our present subject. It is, however, rather interesting. It shows the whole gamut of frequencies with which we are familiar. This plot has near its lower end a frequency of one cycle per 2,000 years which is supposed to characterize a particular comet. From this it proceeds through the frequencies corresponding to solar periodicities, through the frequencies used in commercial power systems, through the voice frequencies, the wire carrier, the radio frequencies, the longer heat waves, the visual light rays, ultra-violet rays, X-rays and to the very hard rays sometimes called cosmic rays with which Dr. Millikin's name is here associated because of the investigations which he has carried out regarding them. This whole matter of frequency range and the relation of each part to human needs is of the greatest significance and interest.

In considering now how these long and short radio waves are handled in forming the transoceanic circuit, we will look first at the transmitting stations and antennæ, next at what happens to these waves in space and then at the receiving antennæ and stations.

At the transmitting end Fig. 5 shows a picture of the long wave antenna at Rocky Point, which well suggests the characteristics of these long waves. A frequency around 60,000 cycles corresponds to a wave length of about three miles and needs these physically large structures to effectively radiate the power into space. This antenna has six towers each 400 feet high. These long waves are in a frequency range which is much used and relatively narrow so that it is essential that the frequencies be employed the most economical way possible. This has resulted in the employment for the long waves of what is known as a single side band carrier suppression method of transmission, a refinement of transmission which, so far as I know, is not employed anywhere else in radio services although it is employed to a large extent in carrier over our wire circuits. With ordinary radio telephone transmission such as is used in broadcasting there is a constant steady frequency emitted even when no speech is being sent out and somewhat over 2/3 of the total energy radiated is in this steady carrier frequency which, of course, transmits no message. In the

system here employed, however, this steady frequency is practically eliminated so that when there is no speech to be transmitted practically no energy whatever leaves this antenna. Furthermore, in ordinary



Fig. 5—A picture of Rocky Point sending antenna

transmission when there is, say, a 1,000-cycle tone to the voice, this appears in the transmission as two frequencies 1,000 cycles above and 1,000 cycles below the carrier frequency. This evidently is an ineffective use of the available frequencies so that one of the so-called frequency side bands in this system is eliminated. In this way a single tone in the voice is represented by a single frequency in the transmission from the station. A further frequency economy by the use of the same frequency range for talking in two directions is brought out below.

Fig. 6 shows the large vacuum tubes used in the last stage of the Rocky Point long wave radio transmitter. As many as 35 such tubes are employed in parallel capable of putting into the antenna something

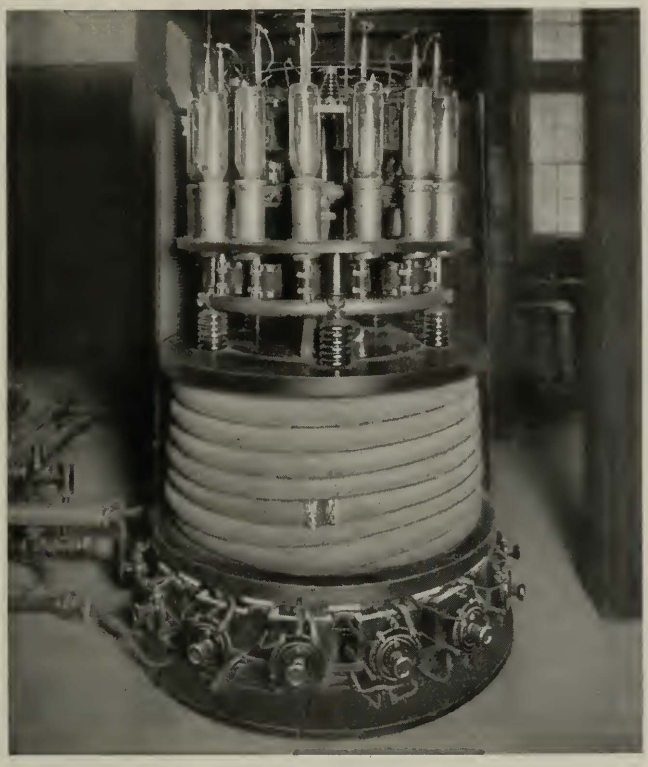


Fig. 6—Showing last power stage

over 200 killowatts which, as already noted, is worth something over six times this amount in the form of ordinary radio transmission.

This is the only picture I shall show of any of the apparatus involved in this work although such apparatus represents, of course, a tremendous amount of fundamental investigation and development and design. In a picture of apparatus, however, you can see little but assembly of cases and wiring and occasional vacuum tubes, and this gives you no adequate idea of what is going on electrically inside of the devices. An interesting feature of the transmitting apparatus in this system is that in the process of suppressing the carrier and stripping off one of the side bands, a double frequency transformation is required. For example, a 1,000-cycle tone coming into the station appears first

as 32,200 cycles and next as 59,500 cycles, which is the frequency transmitted.

Fig. 7 gives us in contrast one of the short wave sending antennæ. This particular one is for a wave of 22 meters wave-length, that is,

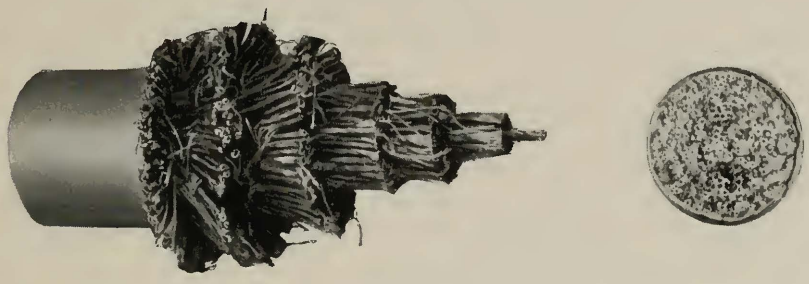


Fig. 7—Deal Beach sending antennæ

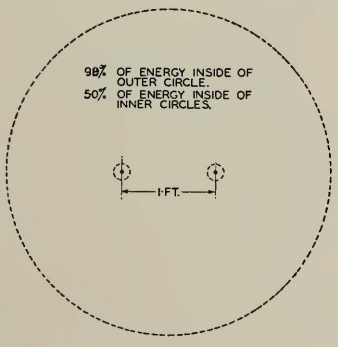
about 70 feet, corresponding to a frequency of around 14,000,000 cycles. These shorter wave-lengths lend themselves readily to directional sending and the resultant possibilities of conserving power in that way. The form of antenna shown is essentially a group of vertical antennæ with a transmission line connecting at points of equal phase. With this particular antenna, energy received in England is increased by

about ten times as compared to that received from a non-directive system.

Assuming we have put the power into proper frequency form and radiated it into space, the next question is how does it fare in traversing the great distances before it reaches the receiving points. We can hardly state this as a technical problem since there is nothing the engineer can do to control it. He can find out merely what nature does to such waves and try to arrange his transmitting, and particularly his receiving systems to meet the characteristics of the waves.



Section of New York-Chicago cable



Energy distribution with No. 8 B. W. G. open line circuit carrying A. C. currents

Fig. 8—Cross-sections of wire line with energy circles and cross-section of cable

We could think of no slide to show this space transmission unless possibly a picture of the world rotating in space such as we have seen adorning popular articles on radio. We will suggest the matter, however, by contrast with wire transmission.

The left-hand part of Fig. 8 shows a cross-section of a pair of copper wires spaced a foot apart on a pole line, which is the standard telephone wire arrangement. On such a circuit 98 per cent of the energy is transmitted inside of the outer dotted circle. The right-hand part of the slide shows two views (one a cross-section) of a typical telephone toll cable of somewhat under three inches diameter. Practically all

the energy for about 300 telephone circuits is transmitted inside this sheath.

While both the radio and the wire transmission involve similar electromagnetic waves, there could hardly be a greater contrast in the method of handling waves than that between the radio transmission we are considering in this paper and transmission employing such wire methods and spanning the comparable distance of say San Francisco to New York.

Recently in visiting the short wave receiving station in New Jersey I was shown oscillographs taken on radio telegraph transmission in which each telegraph dot was followed about a tenth of a second later by what appeared to be an echo. The first transmission came some 3,000 miles from England. The second transmission had gone the opposite direction around the world and had travelled some 22,000 miles before reaching the same receiving point. In such long distance radio we may then have a situation in which each individual signal sets up oscillations, perhaps measurable oscillations, in space surrounding practically the whole earth.

Contrast this to the toll cable shown in the picture which, as already noted, contains about 300 circuits. Such cable is used now commercially for distances up to around 1,500 miles and is permissible for 3,000-mile distances such as we are here considering. In such cables each message is practically confined to a strip of space extending between the terminals of the cable and smaller around than a lead pencil. In the radio case we have literally all out of doors but there is little we can do to control it. In the cable case the channel is reduced to the meagerest dimensions but this so reduced space we can pretty nearly call our own, surrounded and shielded as it is by a sheath and containing carefully balanced circuits. Such space is only occasionally penetrated by outside disturbances when some of our power friends set up unusually strong electrical fields in its immediate neighborhood. By loading, by amplifiers, by equalization of various sorts, this meager space is guarded and controlled and rendered efficient and constant.

We are still somewhat in the dark as to how kind nature has been to us regarding short waves and what degree of reliability we can ultimately get from a circuit employing them. There is nothing yet in the picture, however, to suggest a reliability for long or short wave radio approaching that of a cable circuit for similar distances.

A large number of measurements have been made of the strength of the radio fields laid down in England from the long and short wave stations in America and similarly in America from the English stations. Along with such data is taken the amount of noise interference present

at the receiving points. Fig. 9 shows data taken in this way. The curve which reaches the highest point shows for a typical summer's day the field strengths received on the long wave from America. The

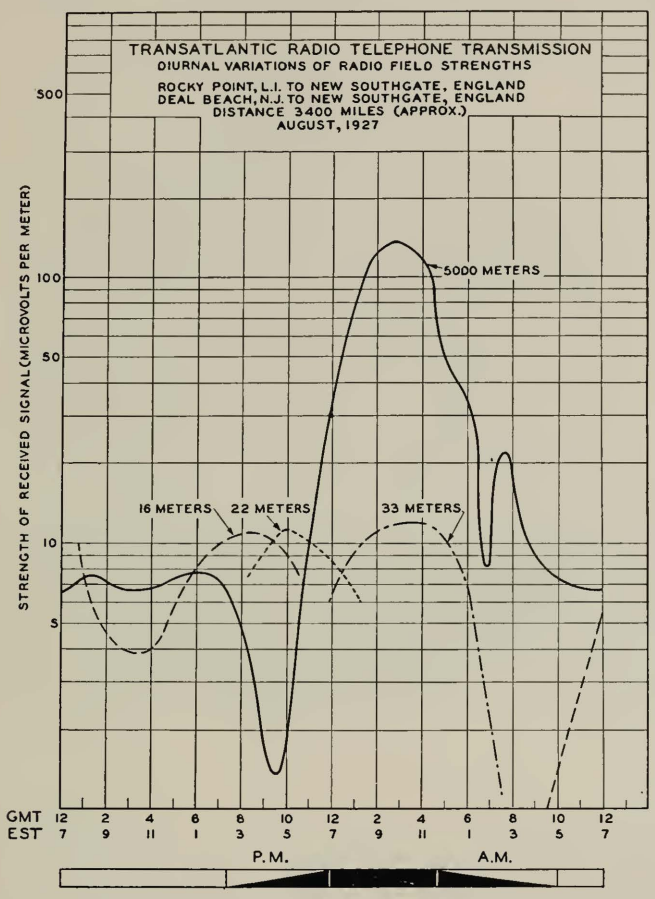


Fig. 9—Long wave curve and corresponding short wave curve of received field strength

other curve shows the field strengths received on short waves employing three wave-lengths, 16, 22 and 38 meters, and taking for each time of day the one of these which was best suited to the conditions.

On this typical day all the wave-lengths were operating well. It will be noted, however, that there are times when the long wave is low and some one of the three short waves is more effective and other times when none of the three short waves are high but the long wave is effective. Furthermore, all of these waves vary a good deal so that

on certain days for hours shown operative on these curves one or more might be entirely out of service. This chart indicates the tremendous advantage in employing a number of separate wave-lengths varying a good deal in their characteristics and choosing at any one time that wave-length which is giving the best performance.



Fig. 10—Showing receiving antenna at Cupar

Having followed the radio transmission as the waves radiate into space and traverse space to the receiving end, we come to the matter of receiving stations. Fig. 10 shows a view of the wave antenna at Cupar, Scotland, used for receiving long waves. This complete antenna arrangement is made up of two pole lines such as shown, each about 3 miles long; a third may be added. These pole lines are placed parallel to each other with separations of about two miles. A pole line joins the two together and connects them to the receiving stations.

It is fortunate that in America (and the same is true to a considerable extent in England) the signals come in from a northerly direction and the static tends to come in from almost the directly opposite direction. The directivity brought about by such antennæ has, therefore, a very large effect. It is estimated that under average conditions the present

antenna gives as much improvement as would an increase in power of about 100 times. Furthermore, receiving at Houlton, Maine, rather than in the vicinity of New York, by getting to a more northerly latitude, is equivalent to a power increase of 50 times. The antenna location and directivity, therefore, is equivalent to a transmitted power increase of 5,000 times. The receiving set employed with these antennæ at Houlton is of a double demodulation type, of very high selectivity.



Fig. 11—Picture of receiving antenna in Cliffwood

Fig. 11 shows a receiving antenna at Cliffwood, N. J., of the type used for short waves. It is constructed of a wooden framework on which are held two parallel sets of conductors made up by bolting together lengths of copper tubing. Possibly you can make out the form of the two sets of conductors. Each set consists of vertical elements a quarter of a wave-length high and a quarter of a wave-length between successive elements. The first element is connected to the second by a conductor connecting the upper ends of the elements, the second is connected to the third by a conductor connecting the lower ends and so on.

The whole effect gives a degree of directivity equivalent to an increase in power of about 15 times. This general question of short wave receiving is a fruitful field for investigation and for the ingenuity of the engineers. A large number of arrangements have been devised

and a good many of them tried. Considerable further work is under way.

At the time when the short waves are in trouble apparently either one of two conditions may obtain. In the first of these there may be considerable field received from the distant transmitting station but this field is made up of the result of transmission over several paths so that the waves received over the different paths react on each other, causing a rapidly fluctuating interference pattern somewhat similar to that well known with light waves.

There are other times, however, when either the field which reaches the receiving point is not of sufficient magnitude to be picked up or is below the static noise level at the particular time.

One interesting fact with these very short waves is that the ignition system in automobiles may create large disturbances and it is important that the receiving points should be kept away from roadways frequented by automobiles. For this and other reasons the New Jersey location will undoubtedly be moved somewhat from Cliffwood.

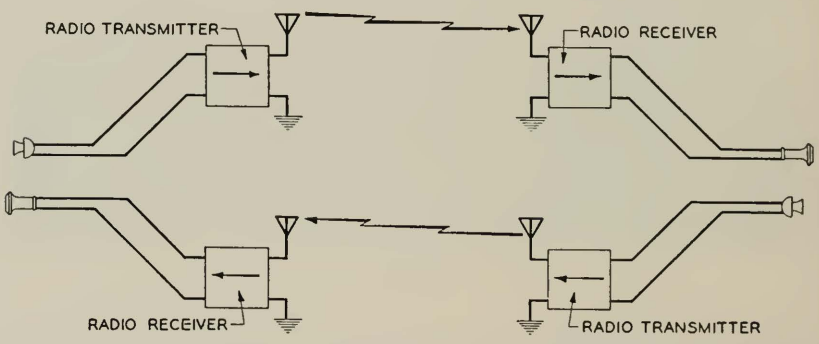


Fig. 12—Indicating diagrammatically east and west transmission on separate channels terminated in ordinary transmitters and receivers

So much then for the question of the radio circuits themselves. The problem remains of making them serve as a link between the two-wire systems on the two sides of the Atlantic. If the problem were merely transmission from one particular subscriber's set to another particular subscriber's set, the very simple arrangement shown in Fig. 12 would be feasible. In this you will note the eastbound and westbound transmission each starting with a telephone transmitter at one end and extending to a telephone receiver at the other are kept entirely separate. Two people could evidently carry on a conversation over this circuit without further complications. In fact this was the way in which the first two-way tests were carried out.

Since eastbound and westbound short wave channels are at entirely different frequencies, there would be no interaction between the east and west-going circuits when used in this way. For the long waves, however, since the east and westbound circuits are at the same frequencies, each transmitting station would send considerable energy into the receiving station on the same side of the ocean, thus giving to each subscriber a heavy side tone of his own speech. This effect, however, could be considerably reduced by separating the transmitting and receiving stations and arranging the directive antennæ of each receiving station so it would receive as little as possible of the corresponding transmitting station. Even with the long waves, therefore, a conversation could be held in this simple way.

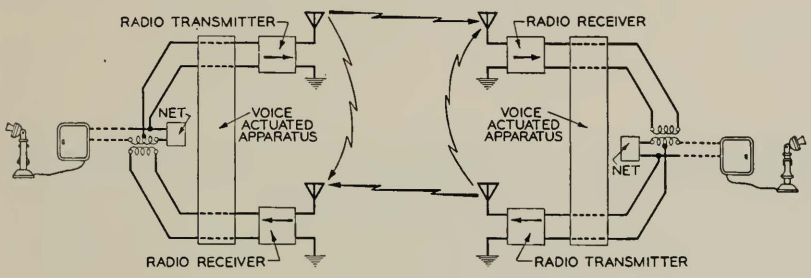


Fig. 13—Transoceanic circuit schematic

When the east and westbound radio circuits are brought together, however, for a connection to a wire circuit at each end, we have introduced some very serious difficulties. Consider first the simpler case of the short waves where the eastbound and westbound transmission are at different frequencies.

The voice waves reaching London from an American talker will be reflected in part either at the London office where the east and westbound channels are brought together or at some point before reaching the European subscriber. Unless means are taken to prevent it, this reflected energy can then pass to the English transmitting station and be transmitted back to America. At the American end a similar partial reflection can take place throwing part of the energy back again to England. In this way, according to circuit conditions, it is possible for the whole circuit either to build up and act as a widely flung oscillator or if the damping at the moment makes this impossible, the speaker and the listener can be much interfered with by electrical echoes and distortion.

In the case of long waves there is the added difficulty as already noted that each transmitting station can throw a good deal of power into the

receiving station on its own side of the ocean, thus bringing in the possibility of local oscillations and distortions.

For either the long or short waves then it is very advantageous, in fact practically necessary, to employ switching devices actuated by the voice waves themselves to prevent the effects just stated. In this diagram you will note drawn so as to involve the wire connections both to the transmitting and the receiving station at each end a rectangle which is labeled "voice-actuated apparatus." This is so arranged that when there is no transmission on the circuit the wires connecting to each of the transmitting stations are short-circuited, making both transmitting stations inactive. Speech coming in then at one of the ends from a telephone subscriber operates a relay which opens the wire circuit to the transmitting station at his end and at the same time short-circuits the receiving path at the same end.

One interesting phase of this voice-operated switching mechanism is the employment of a delay circuit. At the New York terminal of the circuit, when the voice currents reach it from some distant subscriber and after these voice currents have actuated the switching

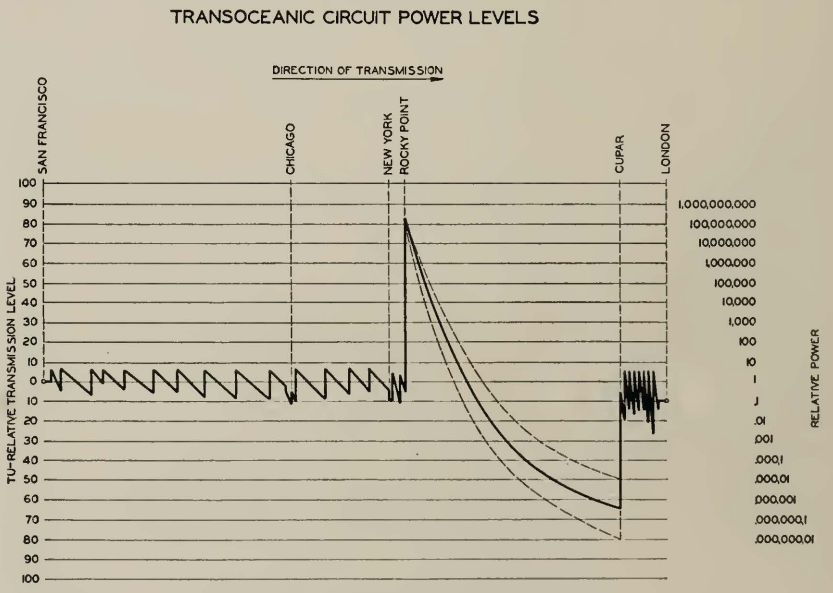


Fig. 14—Diagrams showing energy levels

mechanism noted above, they pass into an artificial line down which they travel, are reflected and travel back to the sending end of the artificial line. In this way the voice waves are allowed to idle away

two one-hundredths of a second during which time the switching mechanisms have performed the operations just noted and have thus put the circuit into shape for the voice waves to go forward.

Fig. 14 shows the shifts in power level in going from one end of the circuit to the other for a connection from San Francisco to London. The zero of the scale corresponds to the power level at which power is given out by an ordinary substation set when actuated by a loud voice. The power ratios compared to this are shown at the right on a logarithmic scale.

The comparatively small ups and downs in the power level corresponding to transmission over the wire lines represent, of course, the line attenuations and successive amplifications by telephone repeaters. The highest power level is naturally at the output of the radio transmitter where it is about one hundred million times the starting level At the English receiving station it has been dropped to about the same ratio below the zero of the scale.



Fig. 15—Technical operators' position

In view of the character of radio, in particular its variable nature, the circuit is under constant supervision during operation by two technical operators, one located in New York and one in London. Fig. 15 shows the special terminal equipment, meters, etc., and the technical operator at the New York end.

It is evidently desirable that the transmitting station shall always put out maximum power whether the speaker has a loud or a weak voice or is near or distant from the transmitting station. One of the duties of the technical operator is to bring this about. By proper indicating meters he knows the power level of the speech going to his transmitting station and he keeps this at the point where it will just completely load the transmitting station.

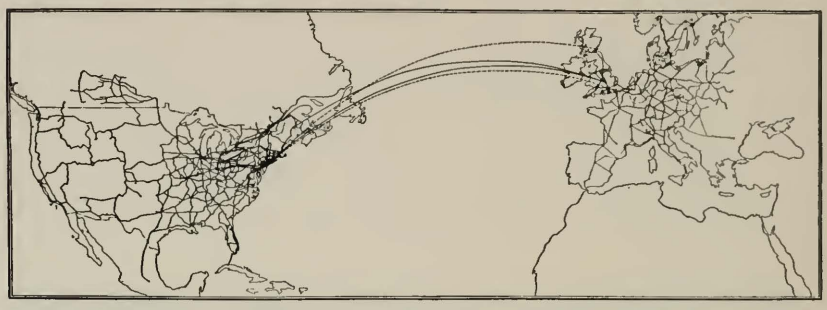


Fig. 16—Fig. 1 with radio connections added

With this brief discussion then let us assume we have progressed from the conditions of our first illustration to the conditions shown here in which the two great telephone systems are joined together by the radio links.

To complete the story, a few more words as to the changes and improvements which development work now under way is expected to give.

As the system stands to-day it does not offer the privacy which ordinary wire connections offer. While ordinary broadcasting sets are not of the proper type to receive messages from this system, it is comparatively easy, with sets designed for the purpose, to pick up one side of the conversation over the system by listening to the transmitting station in the same country as the listener. Because of the voice-actuated devices, the other side of the conversation has to be picked up directly from the distant country, which is a much more difficult thing to do.

To give this system a high degree of privacy is difficult, particularly with the long waves. The frequency range in which the long waves are situated is, as already stated, narrow and well filled so that any proposition that widens the required frequency range, such for example as shifting the carrier frequency rapidly, cannot be employed. Methods have been developed, however, and equipment is far advanced on a

system that is expected to give conversations over this radio channel a sufficient degree of privacy. Certain features of the new privacy method will probably be in experimental use within a few months. It will be at least six months and possibly a year before the complete privacy system is in full operation.

The feature of this whole transatlantic service which worries the engineer most, however, is the matter of reliability. After the engineer has done the best possible in transmitting and receiving stations he is confronted by the fact that transmission through space and noise conditions vary so much that thousands of times as much transmitting power as would be sufficient under good conditions may be inadequate to get through under poor conditions. His only defense is to use a considerable number of wave-lengths which tend not to get into difficulties at the same times or under the same conditions.

Considering first the short waves we find as already noted there is sufficient difference between the transmission characteristics of different parts of the range, between say 10 meters and 30 meters, so that the reliability is considerably improved by designing the stations to use any one of three or more frequencies in this range.

We shall be very happy if by such use of a number of short wave-lengths and by further improvements in technique the reliability of short wave channels can be made such as to some day eliminate altogether the necessity of the long wave channel with its much more extensive plant. There are a number of projects going ahead in the world for the establishment of long distance transoceanic telephone circuits employing short waves alone. With a reasonable further development of the short wave art such service will undoubtedly prove well worth giving.

Telephone service is, however, necessarily an exacting service, particularly since the subscribers participate directly in each connection. Moreover we are dealing here with the joining of North America and Europe which, commercially and otherwise, is of so large importance as to justify perhaps much more exacting technical requirements than any other transoceanic connection.

So far, the data available regarding the short waves do not suggest that they ever will give a reliability of service comparable to that for similar distances over land wire circuits. It is our present expectation, therefore, that the giving of suitable service between America and Europe will require the continuation of the long waves even though such waves demand a much more extensive and complicated plant than do the short waves. In addition to the long waves we shall also want the very best we can get from the short waves. By the combination

of this one long and several short waves we believe that ultimately the service, except for the three summer months of high static, will be but little interfered with by electrical weather, and that even for these months the service will be operative for better than 90 per cent of the time.

You will understand, of course, that the connection to-day from London will be entirely by long waves, the short waves not being ready for commercial operation. The connection from America to England will also in all probability be by long waves, although the short waves are being held in readiness as an emergency routing.

Transatlantic Telephony—Service and Operating Features

By K. W. WATERSON

Synopsis: This paper describes some of the differences in operating practice on the two sides of the Atlantic and plans which were worked out for taking account of them in the handling of commercial transoceanic calls. Difference in the language is also another problem which has required solution. Data are included giving an idea as to the extent to which the transatlantic connection was used during its first year, there having been established during this time a total of something over 2,300 connections.

THE introduction of telephone communication between Great Britain and the United States required the fitting together of the practices of two telephone organizations. The development of usage between subscribers in the two countries involves questions of different telephone habits and experience. It may help to define the problem of setting up a service of this kind if at the start I mention one or two of the more important characteristics of long distance service in the two countries which illustrate outstanding points of difference.

In Great Britain, only number service is available, that is, a service under which the telephone administration undertakes merely to obtain a connection with a specified telephone and on which the message toll charge is assessed in all cases where an answer is obtained from the telephone called whether or not the person desired is there. In the United States, this same number service is available at about the same initial rates as in England. In addition, we have a so-called person-to-person service on which, for a charge approximately 25 per cent above that for number service at the longer hauls, we undertake to obtain connection with a particular person who is specified in the order for service. In case of inability to reach the particular person desired, the full message charge is not assessed, but a so-called report charge is made, which is about 25 per cent of the station charge, tapering off in percentage to a maximum charge of one dollar. This difference in the class of service available in the two countries is a matter of importance because of the fact that our experience here in America shows that on the longer hauls and at the higher rates about 85 per cent of the calls are on a person basis, whereas at short hauls the large majority of calls are for a number only.

In both countries the toll rates provide for an initial talking period of 3 minutes. In Great Britain, additional use of the line is charged

for on the basis of 3-minute units and for each the charge is the same as the initial rate. In the United States, additional use is charged for on a one-minute basis—each minute's charge being about one-third the initial—a finer measure of actual use and one which, particularly at the higher rates, makes long distance telephoning considerably less expensive.

In the United States, the general practice is to allow subscribers to talk as long as they wish except on rare occasions due to emergencies such as storm breaks. In Great Britain, subscribers are notified at the end of 3 minutes and are limited to a maximum use of 6 minutes if other calls are waiting. This difference in the allowable length of long distance telephone conversations has developed out of basic differences in toll service policy as regards the provision of plant and the resultant speed of service. In the United States, we plan to give a very rapid service and we provide toll line plant to meet these needs. This policy seems to best meet the needs and desires of American users and to have been a large factor in the rapid development of our toll business. As a result, practically all toll and long distance calls are placed at the time when the connection is wanted and subscribers are often impatient if their calls are not completed immediately. In Great Britain, the plan has been to maintain as high efficiency as practicable in toll line plant and this naturally results in a somewhat slower long distance service. British subscribers are accustomed to longer delays than ours in obtaining connections. There is considerable advance booking. Under this condition, the limitation of the talking period, which I have already mentioned, is a practicable means of making the service available to as many users as possible and also of avoiding possible cases of unfair use of the lines by certain individuals to the exclusion of others. This difference in practice regarding the allowable length of conversation is a matter of particular moment in connection with a service like the transatlantic service for the reason that our experience has shown a definite tendency for users to talk for longer periods on the long haul, higher rate business. This is probably indicative of the greater importance or different nature of this class of telephone communication. Whatever the reason, calls at 250 miles average under 5 minutes, at 500 miles—$5\frac{1}{2}$ minutes, at 1,000 miles—6 minutes, and transcontinental calls—$6\frac{1}{2}$ minutes.

In Great Britain, distances are relatively short. London-Glasgow, for example, represents one of the important longer haul routes, and the air-line distance is about 350 miles. On international calls, London to Berlin is one of the longer hauls at which service is available

and this is under 600 miles. In Great Britain, there is relatively small development of telephone usage at these distances. In the United States, on the other hand, we have transcontinental service over some 3,000 miles with considerable business at this and other long distances. Our connection with the Cuban Telephone Company has also given us experience with very long haul service. So in the matter of special long distance problems and in the development of long distance telephone usage, the experience has been largely on this side of the water.

The service arrangements for this transatlantic undertaking were made through discussions carried on in London by representatives of our organization and officials of the British Post Office. While there were a good many problems to be worked out, there was the usual result, when both parties desire to cooperate and to discover the best solution, that an agreement was soon reached. It was decided that the service needs of transatlantic telephony would best be met by a single class of service with one rate covering either number or particular person usage. Experience in the Bell System had indicated that on long haul business of this nature, practically all calls would be for a designated person and this has been borne out in the transatlantic usage. The rate between Great Britain and twelve states in the northeastern part of our country was fixed at $75 for 3 minutes, with an additional minute charge of $25. A report charge, of which I have already spoken, was fixed at $10 for use in certain cases where particular persons called cannot be reached.[1] The British Post Office preferred to apply the same rate to England, Wales and Scotland. Because of its wide expanse and expensive land line plant, the United States was divided into five zones for fixing additional land line charges over and above the New York terminal rate. These rate zone lines follow state lines. The zone rates go up in $3 steps as we draw away from New York. Zone rates follow reasonably well the land line charges for service from New York. This zoning plan was adopted to simplify the means of quoting and computing the transatlantic rates abroad as compared with superimposing the more finely measured land line rates on the New York terminal charge.

For communications extending beyond the initial 3-minute period, the plan of charging on a single-minute basis was adopted, as it seemed the most equitable, particularly in view of the distances and charges involved.

[1] "Reduced rates for transatlantic service became effective March 4th superseding those mentioned in this paper. To illustrate the extent of the reductions, a three minute call from New York to London which was initially $75. is now $45."

In setting up service and operating arrangements, it was necessary to give consideration to different conditions which would exist dependent upon the volume of business to be handled over the transatlantic channel. We had to consider operating practices which could be used satisfactorily either under conditions of high load and possibly delayed service or of light load and fast service. As a means of insuring service to as many users as possible in periods of heavy business, agreement was reached that there should be a 12-minute limit on individual usage in case other calls were awaiting assignment to the radio channel. So far, there has been no occasion to enforce this limitation. The limitation of 12 minutes was adopted instead of the usual 6-minute limitation common in British telephone practice for the reason that due to the relatively long talk periods on business of this kind, the 6-minute limitation would have resulted in interfering with too large a proportion of these communications.

One problem of interest involved in the transatlantic service was that of fixing rates which allowed of satisfactory expression either in terms of English pounds and shillings or in American dollars. For this purpose, 4 shillings was considered the equivalent of an American dollar. The rate from London to New York, for example, is £ 15 for 3 minutes and £ 5 for each additional minute. Our zone rate steps of $3 for the initial period and $1 for each additional minute were so set in order to allow of even dollar and even shilling quotations for the zone charges. The rate from Cleveland, for example, which is in our second zone, to London is $78 for 3 minutes and $26 for each additional minute. The same rate quoted from London is £ 15: 12 s. for 3 minutes and £ 5: 4 s. for each additional minute. Rate treatment of this kind was thought desirable, not only to allow of easy expression of rates in either English or American money, but also to avoid odd cents in our service charges.

Another problem had to do with the fixing of the hours of service so that the service would be most valuable and usable with due regard to the five hours difference in time between New York and London. At the time the service was opened, limitations on the use of the Rugby sending station for telephone transmission made it possible to keep the channel open only $4\frac{1}{2}$ hours during the day. The hours from 8:30 A.M. to 1:00 P.M., New York time, which correspond with 1:30 to 6 o'clock, London time, were adopted as allowing the maximum overlapping of the London and New York business day. Later, it became possible to extend the hours of operation so that now the service is available $10\frac{1}{2}$ hours—7:30 A.M. to 6 P.M., New York time, which is 12:30 to 11 P.M., London time. The fact that

both London and New York are on a daylight saving schedule in the summer months has required some shifting of the hours of service as these time rearrangements are effected on the two sides of the water.

The operating arrangements set up for the handling of this business provide traffic control operation at the New York and London long distance offices. These offices have direct access to the radio channel via the radio stations at Rocky Point and Houlton and at Rugby and Cupar where technical operators have the transatlantic channel under constant supervision and control. The New York and London long distance offices have special equipment arrangements necessary for connecting the radio channel and the land lines. On calls terminal at New York or London, the operation is similar to that on other terminal calls. On calls involving points beyond New York and London, the New York and London operators assume control, holding the land lines in readiness for prompt connection to the transatlantic channel, supervising the connection and fixing the amount of chargeable time, special measures being provided to protect the user from overcharges that might result from conversations being longer than otherwise necessary because of static and other atmospheric disturbances. The operating method is set up to require a minimum of time on the transatlantic channel for passing calls back and forth and preparing connections.

The personnel necessary to operate the transatlantic circuit is probably not generally appreciated. While two operators in London and two in New York can readily handle the calls themselves, there are six stations, three in each country, for operating and controlling the radio channel and from 35 to 40 men are needed for this work. This force could, of course, handle much more business than is now offered.

Just as the experience with special long distance operating problems and with the development of long distance usage had been largely on this side of the water, so in the matter of international telephone arrangements the experience had been largely with the British Post Office. We have had connection with Canada for many years, but in none of our interchange of business with Canadian companies have we encountered the problems incident to European international communication. The British Post Office, on the other hand, has communication with many countries on the continent and has played its part in the various European conventions and conferences looking to the betterment of international telephone agreements and communication in Europe. So their experience was particularly helpful in shaping up the contract arrangements. In general, the contract

between the British Post Office and the American Telephone and Telegraph Company covers such matters as responsibilities of the two administrations, classes of service, rates, broader operating provisions and settlement matters.

Turning now to the question of the results which are being obtained in this transatlantic service. During the first year, something over 2,300 connections were established. This is an average of about 7 a day, if we include Saturdays, Sundays and holidays, on which days as a general rule the flow of telephone traffic is relatively low. Usage is not very different east and west, something like 55 per cent of the business having originated on this side. Some business from the other side is, of course, from traveling Americans. After the first two months, January and February, when the business amounted to about 250 messages a month and was affected largely by formal openings and curiosity calls, the traffic fell off, and during the summer it was not more than half as great as it had been in the first two months. This may have been due partly to falling off in business activities and possibly also partly to the fact that more atmospheric difficulties are experienced during the summer and the service is then somewhat less dependable. As a matter of fact, there was less atmospheric difficulty than we had anticipated. Starting with September, the business has shown a steady increase. On Christmas Day there were 44 messages.

About half of the transatlantic calls are between New York and London. Over 70 per cent of them originate or terminate in New York City and the remaining calls involve points scattered over the rest of the country. Considering the type of usage of this transatlantic service, nearly half of the calls appear to be of a social nature. As to calls for business purposes, banks and brokerage concerns account for the greatest use so far.

In general, the quality of speech transmission has been more satisfactory than the preliminary tests indicated it would be possible to maintain throughout the year. The radio link is, of course, under careful observation throughout the service period and is not assigned for commercial use unless it appears that reasonably good communication will be obtained. Except for two summer months when atmospheric conditions made telephone communication impossible on an average of about 2 hours a day, the lost time due to static and other such troubles in the radio channel has been relatively small.

Except for brief periods on individual days, the traffic volume has not been sufficiently high to result in any problem in providing a fairly prompt service. At times, and particularly during the summer

months, individual calls have been delayed due to the fact that at the time they were offered, atmospheric conditions made it impossible to use the transatlantic channel. As the business develops, it will doubtless be necessary to adopt special measures for evening the flow of business throughout the period that the transatlantic channel is open for service. At such time as traffic develops to a point where some artificial leveling of the load is required, we would expect this service to involve advance bookings and longer delays than we are accustomed to here in the United States in our internal services. Pending the availability of other transatlantic channels through the use of short wave lengths or otherwise, I do not believe that this type of service would necessarily be seriously objectionable or deterrent to business development. As a matter of fact, a good many calls are now filed in advance.

Differences in the English language as spoken in London and New York became evident as soon as our New York operators were placed in communication with the operators in London. Each group expressed some concern as to what the other was doing to their language. I believe the London operators were inclined to think the broken English spoken by the telephone operators in Holland was sometimes easier to understand than New York City English.

The self-confidence of Americans evidenced itself in the considerable number of calls filed for the nobility, cabinet ministers and other men in the public eye. The fact that most of these calls were accepted by the persons called, indicated their willingness to play the game. Other evidences of this same confident attitude were the suggestions from individuals that they be given a free call so that we could capitalize on the publicity that they would put into their advertising. Others advised us after using the service that, for a consideration, they would allow their names to be used in our publicity material.

I have spoken of some of the operating problems in setting up the transatlantic service and of our experience in handling this service since its inauguration about a year ago. The operating and service arrangements have worked out satisfactorily. The service as a whole has been considerably better than we had anticipated. The volume of business is small but the business now being handled is in line with our general experience in the development of long distance usage. In a situation of this kind, full consideration should be given to the fact that, generally speaking, potential traffic volumes decrease with distance. Telephone service like the transatlantic is a new means of communication. It will not only take time for potential users to become convinced that satisfactory communication can be carried on

by telephone but time is required before they will break away from dependence on other means of communication such as the cables and mails with which they have had long experience and which may have appeared to meet their needs.

The development of our transcontinental business is of interest as this route may be considered as close a parallel as we can find to the transatlantic situation. For several years after the opening of the transcontinental service, the business was small but it has since greatly increased.

The transatlantic channel is a radio channel and this suggests a possible lack of privacy such as is obtained in ordinary telephone communication. Actually, the chances for conversations being picked up by persons to whom they would be of interest or value are rather remote, but this possibility has doubtless had some deterrent effect on the development of business. It is expected that these deterrent factors will be removed in the near future through the introduction of new equipment arrangements which will assure a high degree of privacy on these overseas conversations.

Possibilities for growth must be present in a communication system at the terminals of which we have New York and London, the largest business centers in the world, both English-speaking. On this side, the service has already been extended beyond the United States to Canada and Cuba, and will be extended to Mexico. On the other side the service has recently been extended beyond England, Wales and Scotland to important cities in Belgium, Holland, Germany and Sweden. Further extensions are under consideration to other important continental cities between which and this country there is undoubtedly potential business. As the service is extended beyond Great Britain, a language problem appears. So far about 5 per cent of the conversations are not in English. For the time being, we are relying on the London operators for smoothing out language difficulties in establishing connections and the problem is, of course, not a new one to them. We are planning, however, to set up an operating force here in New York which can communicate in their own language with users who are not speaking in English.

Not only from a technical viewpoint but in other respects we are gratified at the results of the first year's operation of the transatlantic service and we look forward with confidence that this service will be, not only the quickest, but an essential factor in communication between the old world and the new.

Phase Distortion and Phase Distortion Correction

By SALLIE PERO MEAD

SYNOPSIS: The importance of the rôle played by the steady state phase characteristics of long cable circuits has recently been emphasized in telephone and telegraph transmission. In this paper an analytical exposition of the theory of phase distortion is followed by a consideration of various methods of phase distortion correction with particular reference to terminal phase compensating networks and to the application of the lattice network to the loaded line as a terminal phase corrector.

1. INTRODUCTION

FROM the standpoint of ideal quality, a transmission system must be so designed that the received currents, which represent the transmitted signal, shall be a faithful copy of the corresponding currents which enter the transmission system at the sending end; that is, the transmission system must be distortionless. For relatively short distances the deleterious effects of phase distortion are not appreciable but as the range of transmission is increased a point may be reached where the impairment of quality becomes so serious as to reduce the commercial efficiency of the circuits. This fact was first recognized on long submarine telegraph circuits. With regard to telephony, the importance of the question of the quality of received speech was initially emphasized by the advent of the efficient telephone repeater,[1] making possible the increased length of the modern telephone system. This increased length, involving the necessity for circuits of higher quality, led to the development of improved loading designs.[2]

It has long been recognized as a principle of good telephone or telegraph transmission that the variation of attenuation over the range of speech or signal frequencies should be minimized. With the increased length of circuits, however, this requirement alone was found insufficient to insure good quality and *phase distortion* over the range of essential frequencies was found to play an important rôle. Thus steady state phase characteristics have attained prominence in the engineering of long cable circuits and in the application of this technique to telegraph, telephone and picture transmission.

Distortion in the variation with frequency of the phase difference between the current at the receiving end and that at the sending end of the transmission line, as well as distortion of the amplitude characteristic of the received as compared to the initial current, gives rise

[1] See reference 1.
[2] See reference 2.

to so-called transient effects. That is, the currents, after arriving at the distant end of the communication circuit, require an appreciable time, varying with the impressed frequency, to build up and, under certain conditions, may never build up to anything remotely resembling the transmitted currents in the short interval during which the latter exist. This effect, moreover, may produce serious impairment in the quality of the received speech or signal even when the line is so designed that the steady state attenuation of all currents in the essential range is substantially constant.

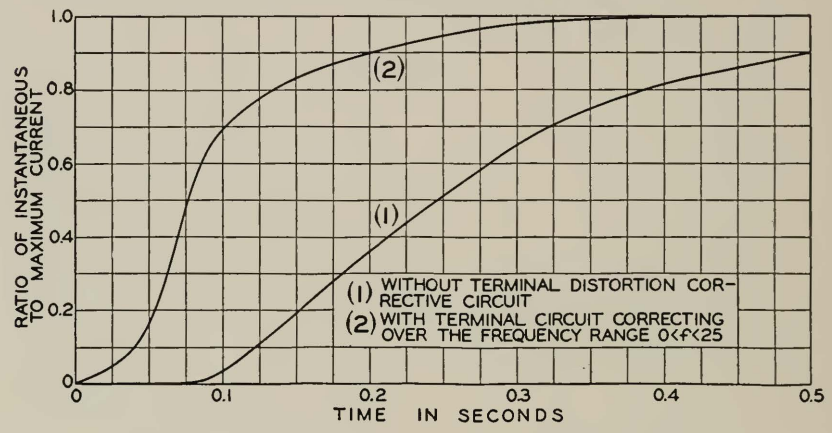


Fig. 1—Building-up of current on 1500 mile telegraph cable

As an illustration of the transient distortion on a transmission line, consider the indicial admittance, $A(t)$, of the cable of length l miles, and of resistance R and capacity C per mile; that is, the received current in response to a unit d.c. voltage applied at the sending end at time $t = 0$. This is given by [3]

$$A(t) = \frac{2}{Rl} \frac{e^{-1/y}}{\sqrt{y}},$$

where

$$y = \frac{4t}{RCl^2}.$$

Curve (1) of Fig. 1 represents relative values of $A(t)$ on a cable 1,500 miles long. (This is the same cable whose phase characteristic is shown in Fig. 5, the inductance being ignorable in determining the indicial admittance.) The departure of the received current from the abrupt wave front of the impressed d.c. voltage is clear.

[3] See reference 3.

The distortion on the loaded cable from the transient point of view is represented in Fig. 2, which shows the envelope of the building-up of current of frequency $\omega/2\pi$ in the Nth section of a periodically loaded line.[4]

The oscillograms of Fig. 3 show the received current in response to sinusoidal waves on a 600-mile medium heavy loaded (H-174) line.[5]

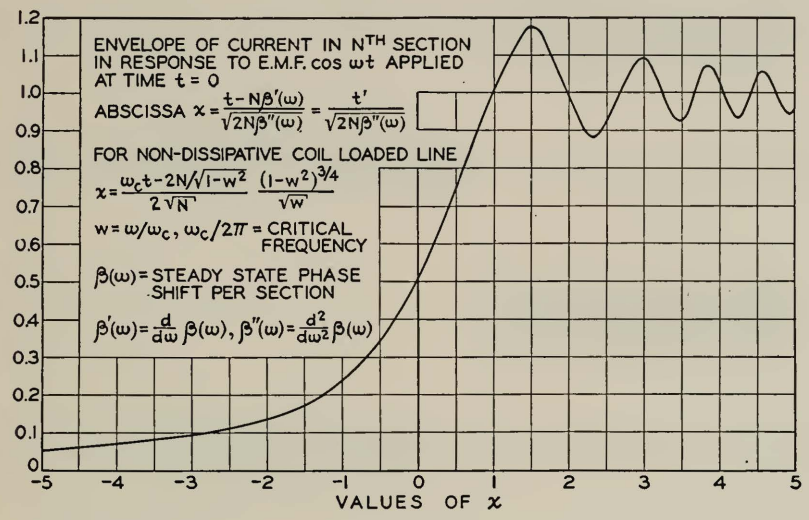


Fig. 2—Building-up of alternating currents in long periodically loaded line

Figs. 3*a* and 3*b* are for frequencies of 1,000 cycles and 1,500 cycles respectively and Fig. 3*c* is for a compound wave made up of 800 and 1,600 cycles. The last oscillogram shows clearly the greater delay of the higher frequency. Due to the relative weakness of this component the building-up transients while quite apparent are not pronounced. It will be observed that an appreciable time has elapsed in each case before the received wave has built up to the amplitude or frequency of the steady state.

The detrimental effects of transient distortion were first studied in the long submarine cable. Early in 1918 a theory of distortion correction was developed by John R. Carson from the transient point of view. The principle was arrived at that the *modified arrival curve* of prescribed form (a square-topped wave in the case of long line telegraphy) may be obtained by combining derivatives with respect to time of the *datum arrival curve* (the current obtained at the end of

[4] See reference 4.

[5] In this symbolism "H" refers to coil spacing of 6,000 ft., and the following number gives the inductance in millihenrys.

the cable itself) in the proper proportions to insure the steepness of building-up of the arrival curve and the maintaining of the tail of the wave. Terminal corrective networks [6] in combination with vacuum tubes were designed to obtain the requisite derivatives. Such a terminal corrective network is shown in Fig. 4, where the

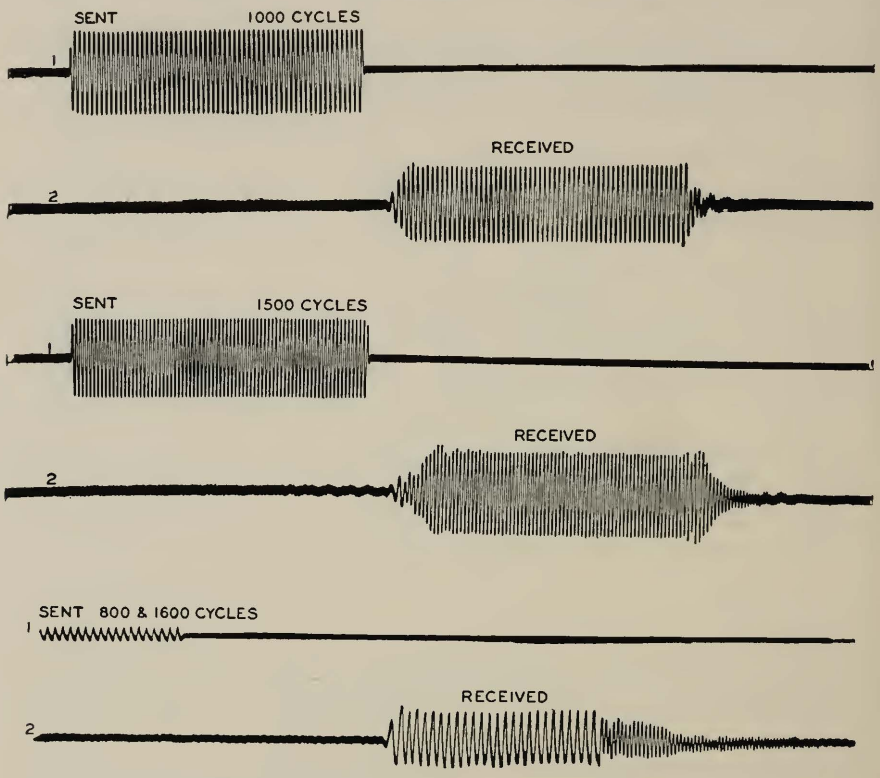


Fig. 3—Transient distortion in a 600 mile length of H-174 loaded cable

resistance R_1 is a distortionless one-way thermionic tube and $V(\omega)$ the output voltage. Examination of this network in the light of the more recent study of phase distortion correction from the steady state point of view has shown that it does also correct the phase distortion of the cable and provide some attenuation equalization as well.

Although the design of corrective networks on the basis of the steady state phase has usually been found more simple than on the transient basis, a knowledge of the arrival current or voltage as an

[6] See references 5 and 6. Also, for the development of distortion corrective circuits with vacuum tube amplifiers, or "signal shaping vacuum tube amplifiers" as they are called, in connection with their application to the new permalloy cables of the North Atlantic, see references 7 and 8.

explicit time function is sometimes essential, in the last analysis, to indicate the degree of correction afforded. This information may be supplied to sufficient precision by the first and succeeding derivatives with respect to frequency of the steady state phase, which, as explained

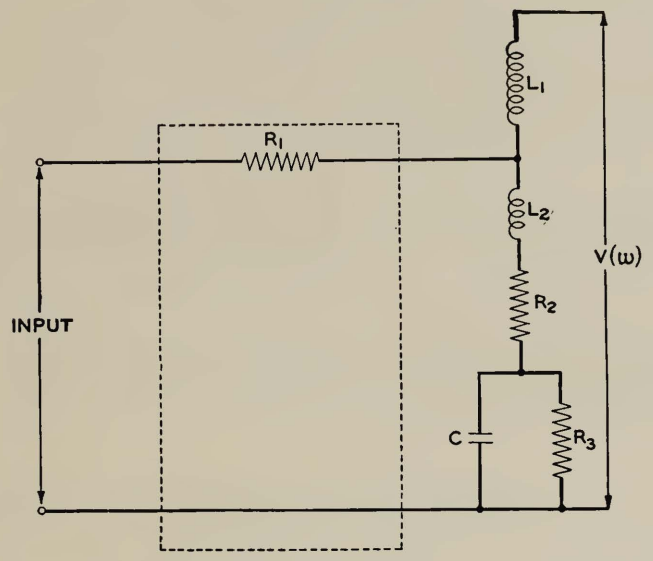


Fig. 4—Distortion corrective circuit for long telegraph cable

below, are extremely useful criteria of the improvement in the time and steepness, respectively, of building-up over a finite range of frequencies. Nevertheless, to visualize the actual effect of the applied phase compensation upon the arrival current or voltage due to an impressed e.m.f. of a specific frequency requires the evaluation of the explicit time function.

It is the object of this paper, following an analytical exposition of the theory of phase distortion, to consider various methods of phase distortion correction with particular reference to terminal phase compensating networks and the application of the lattice network to the loaded line as a terminal phase corrector.

II. Phase Distortion

1. *Steady State Theory of Phase Distortion*

The mathematical theory of phase distortion in signaling systems may be explained briefly from the steady state point of view. We suppose that it is required to transmit a signal $f(t)$, which can be represented by the Fourier integral,

$$f(t) = \frac{1}{\pi}\int_0^\infty F(\omega)\cos\,[\omega t + \theta(\omega)]d\omega, \tag{1}$$

and that the transfer impedance of the transmission system is given by

$$Z(i\omega) = |Z(i\omega)|e^{iB(\omega)},$$

where $\omega/2\pi$ is the frequency. The received current is, then,

$$I(t) = \frac{1}{\pi}\int_0^\infty \frac{F(\omega)}{|Z(i\omega)|}\cos\,[\omega t + \theta(\omega) - B(\omega)]d\omega. \tag{2}$$

$F(\omega)$ exists for all values of ω from zero to infinity but, practically, $F(\omega)$ is negligible except over a finite range which is determined by the nature of the signal. For program transmission, for example, the essential frequencies are now considered to lie in a band from about 100 to 5,000 cycles, while for slow speed telegraphy they lie in a band between zero and 10 or 20 cycles per second. If we suppose, then, that the essential frequency band extends from $\omega_1/2\pi$ to $\omega_2/2\pi$, we may replace equations (1) and (2) by

$$f(t) = \frac{1}{\pi}\int_{\omega_1}^{\omega_2} F(\omega)\cos\,[\omega t + \theta(\omega)]d\omega \tag{3}$$

and

$$I(t) = \frac{1}{\pi}\int_{\omega_1}^{\omega_2} \frac{F(\omega)}{|Z(i\omega)|}\cos\,[\omega t + \theta(\omega) - B(\omega)]d\omega. \tag{4}$$

Now suppose that within the band of essential frequencies, $\omega_1 < \omega < \omega_2$, we have

$$|Z(i\omega)| = R \tag{5}$$

and

$$B(\omega) = \omega\tau \pm n\pi,$$

where R and τ are constants and $n = 0, 1, 2, \cdots$. Then we may write

$$I(t) = \pm\frac{1}{\pi R}\int_{\omega_1}^{\omega_2} F(\omega)\cos\,[\omega(t-\tau) + \theta(\omega)]d\omega, \tag{6}$$

$$= \pm\frac{1}{R}f(t-\tau),$$

$I(t)$ being positive or negative according to whether n is even or odd. Whence the received current is proportional in amplitude to the applied signal and merely delayed in time by the 'transmission time' τ. Thus the received current has the same wave form as the applied

signal or the transmission is distortionless. Accordingly, we have the following proposition.[7] *The necessary and sufficient condition for the practically distortionless transmission of signals in communication systems is that, over the essential range of frequencies contained in the transmitted signal, the transfer impedance of the transmission circuit be equalized both as regards amplitude and phase; that is, the amplitude must be constant and the phase angle linear in the frequency, with a value, when the frequency is zero, of* $\pm n\pi$, *where* $n = 0, 1, 2, \cdots$.

For many years the variation of the phase angle with frequency was ignored. Research in distortion correction was directed to devising networks [8] so designed that $|Z(i\omega)|$ would be a constant, R, over the range of essential frequencies. Assuming that this condition is fulfilled by the transducer but that

$$B(\omega) = \omega\tau + \sigma(\omega) \pm n\pi,$$

where $\sigma(\omega)$ is non-linear in the frequency, we may write (4) as

$$I(t) = \pm \frac{1}{\pi R} \int_{\omega_1}^{\omega_2} F(\omega) \cos [\omega(t - \tau) + \theta(\omega) - \sigma(\omega)] d\omega. \qquad (7)$$

In formula (7) the *amplitudes* of the component frequencies of the arrival curve are, within a constant, the same as those in the impressed signal $f(t)$. The *wave form* of the arrival curve, owing to the presence of the phase $\sigma(\omega)$, may, however, be widely different from that of the impressed signal.[9]

2. *Examples of Phase Distortion in Transmission Systems*

Let us consider the frequency-phase angle characteristic of the two important transmission systems, the submarine telegraph cable and the loaded line.

The cable of characteristic impedance $k = \sqrt{(R + i\omega L)/i\omega C}$ and propagation constant $\gamma = \sqrt{(R + i\omega L) i\omega C}$ (with negligible leakage) is assumed terminated in its characteristic impedance at $x = l$ so that reflection is suppressed. The transfer impedance $Z(i\omega)$ is then

$$\begin{aligned} Z(i\omega) &= ke^{\gamma l} \\ &= |Z(i\omega)| e^{iB(\omega)}, \end{aligned} \qquad (8)$$

[7] See reference 9.
[8] See reference 10.
[9] In telephone transmission it is not at all certain that preservation of wave form is essential. It is essential, however, that the components of different frequencies build up at approximately the same time. It is further demonstrated in the section on 'Loading Systems' below that $\sigma(\omega) = 0$ is the necessary and sufficient condition to fulfill the latter requirement.

where $B(\omega) = \omega\tau + \sigma(\omega) \pm n\pi$ and is the phase angle of the transfer impedance, provided, as we shall assume, that k is approximately a constant. Whether k is a constant or not, $B(\omega)$ represents the difference in phase between the currents at the sending and receiving ends. $B(\omega)$ is given by

$$B(\omega) = l\omega\sqrt{LC}\sqrt{\tfrac{1}{2}[1 + \sqrt{1 + (R/\omega L)^2}]}. \tag{9}$$

Fig. 5 shows $B(\omega)$ and $\sigma(\omega)$ in radians for a 500-mile length of cable whose constants are:

resistance $R = 2.74$ ohms per mile,

inductance $L = 0.001$ henry per mile,

capacity $C = 0.296$ microfarad per mile,

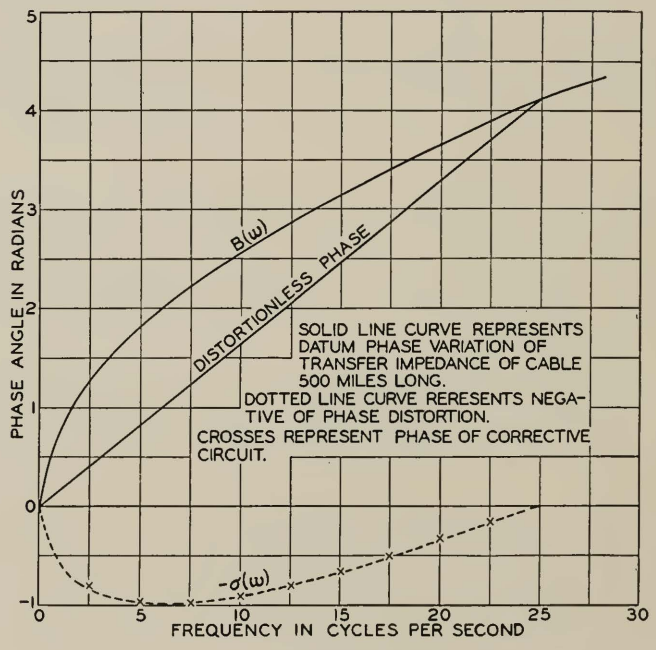


Fig. 5—Phase distortion correction on long telegraph cable

and for the frequency range, 0–25 cycles per second. The straight line, $\omega\tau \pm n\pi$, representing the distortionless phase characteristic is not fixed except that it must pass through the origin or $\pm n\pi$ at zero frequency. It is here chosen to pass through the origin and to have the same value as the cable phase itself at $f = 25$ c.p.s. The dotted curve representing $-\sigma(\omega)$ then shows the departure (with the sign reversed) of the cable phase from the distortionless characteristic.

With regard to the loaded cable, we have similarly the phase angle $B(\omega)$ of the transfer impedance of the cable of length l miles with load impedance Z and smooth line constants R, L, G and C per mile, given rigorously by

$$B(\omega) = \text{imag. comp.} \frac{l}{s} \cosh^{-1} \left[\cosh \gamma s + \frac{Z}{2k} \sinh \gamma s \right], \qquad (10)$$

where

s = spacing of load coils in miles,

$\gamma = \sqrt{(R + i\omega L)(G + i\omega C)}$,

$k = \sqrt{(R + i\omega L)/(G + i\omega C)}$.

It has been found, however, that dissipation and the distributed nature of the line constants have no appreciable effect on the phase

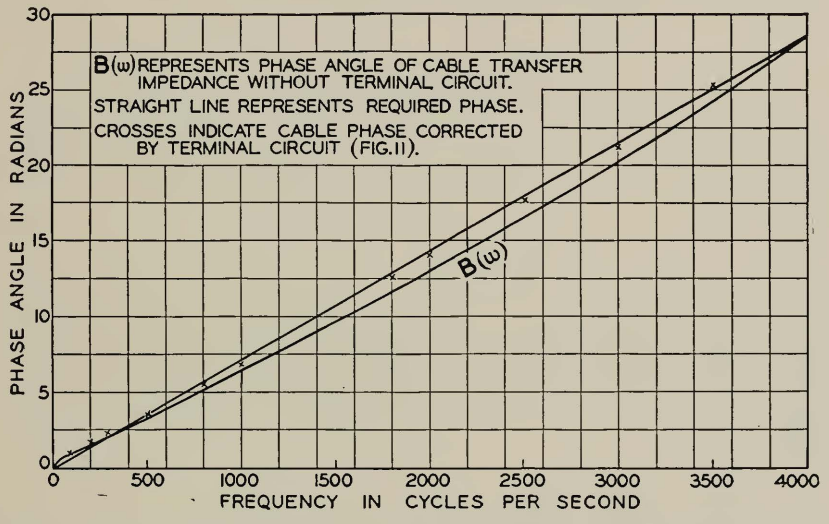


Fig. 6—Phase distortion correction on 20 mile 19 gauge H-44-25 loaded cable, $0 < f < 4000$

variation below 4,000 cycles. Hence, as a close approximation, we may take $B(\omega)$ for the non-dissipative cable without distributed inductance; i.e., simply,

$$B(\omega) = 2N \sin^{-1} f/f_c, \qquad (11)$$

where

$$f_c = \frac{1}{\pi\sqrt{L_0 C_0}},$$

N = number of sections,

L_0 is the coil inductance and C_0 the lumped line capacity per secti on

Even on the light loaded lines designed especially for good quality on long repeatered circuits, the phase distortion is appreciable. The nominal cut-off of these circuits is about 5,600. Fig. 6 shows the phase characteristic of the transfer impedance of a section of side circuit of 19 Gauge H-44 cable only 20 miles long. On Fig. 10 is represented the negative of the phase distortion, $\sigma(\omega)$, obtained by taking $n = 0$ and $\tau = B(\omega_m)/\omega_m$ where $\omega_m/2\pi$ is taken as the highest essential frequency, in this case 4,000 cycles.

In speaking of a pure sinusoidal wave of only one frequency, a phase shift of more than 2π radians or one cycle would be meaningless since every cycle is identical to the preceding and the following cycles. To consider the variation of phase shift over a range of frequencies, however, the total phase shift at any frequency as compared to that at the lowest frequency of the range is required.

III. Phase Distortion Correction

1. *Terminal Networks: Application to the Submarine Cable*

The device of a terminal network having a compensating phase distortion, that is, a network having the phase angle of transfer impedance,

$$\phi(\omega) = [\omega\tau' - \sigma(\omega) \pm n\pi], \tag{12}$$

over the frequency interval $\omega_1 < \omega < \omega_2$ (τ' being a constant), is theoretically the most simple and, in practice, is probably the most flexible and effective method of phase distortion correction. Such a distortion corrective network, in series combination with the transducer in which the attenuation has been equalized, produces an arrival curve

$$I(t) = \frac{1}{\pi R}\int_{\omega_1}^{\omega_2} F(\omega)\cos[\omega(t - \tau - \tau') + \theta(\omega) \pm n\pi]d\omega, \tag{13}$$

which is proportional to

$$f(t - \tau - \tau')$$

provided, of course, that there is no reflection at the transducer terminals. The constant phase angle $\pm n\pi$ does not affect the sinusoidal wave form but merely changes the sign of the wave if n is odd.

The terminal phase corrective network or phase compensator is applicable, at least theoretically, to any type of phase distortion correction and may be supplementary to other forms of correction

such as loading, for instance. Fig. 7 is a schematic diagram of the arrangement of the given transducer of transfer impedance $Z(i\omega)$ with phase angle $B(\omega)$ and the terminal phase distortion corrective network of transfer impedance $N(i\omega)$ with phase angle $\phi(\omega)$. In

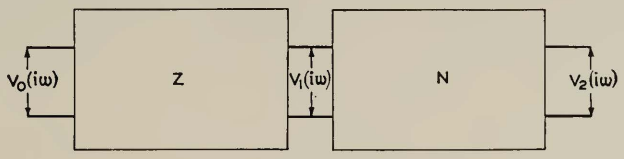


Fig. 7

response to the impressed voltage $V_0(i\omega)$, the voltage $V_1(i\omega)$ at the output terminals of the transducer, which is assumed proportional to the current, is then

$$V_1(i\omega) = \frac{1}{|Z(i\omega)|} e^{-iB(\omega)} V_0(i\omega)$$

and the final voltage $V_2(i\omega)$ is

$$V_2(i\omega) = \frac{1}{|N(i\omega)|} e^{-i\phi(\omega)} V_1(i\omega).$$

Thus

$$\frac{V_2(i\omega)}{V_0(i\omega)} = \frac{1}{|Z(i\omega)|} \frac{1}{|N(i\omega)|} e^{-i[B(\omega)+\phi(\omega)]}. \tag{14}$$

In practical applications, it is usually found advisable to take both τ' and n of equation (12) equal to zero. Then the required phase characteristic, $\phi(\omega)$, of the transfer impedance of the corrective network is

$$\phi(\omega) = -\sigma(\omega).$$

The function $-\sigma(\omega)$ is drawn in Fig. 5 for the submarine cable where $0 < \omega < \omega_m$ and $\omega_m/2\pi = 25$ cycles per second. This may be represented quite closely analytically by the expression

$$\phi(\omega) = \tan^{-1} \frac{ax}{1 + bx^2}, \tag{15}$$

where

$$x = \frac{\omega}{\omega_m} \left(1 - \frac{\omega_m^2}{\omega^2} \right). \tag{16}$$

Thus

when $\omega = 0$, $x = -\infty$ and $\phi = 0$,

when $\omega = \omega_m$, $x = 0$ and $\phi = 0$,

and

when $0 < \omega < \omega_m$, $-\infty < x < 0$ and $\phi < 0$.

Physically, this is realizable in the circuit of Fig. 8 consisting of a resonant element L, C, where $\omega_m = 1/\sqrt{LC}$, in parallel with a resistance R_1. The final voltage is taken across the resistance R_2 and the resistance r represents a vacuum tube. This unilateral element permits of suitable amplification and prevents reflection at the cable terminals so that the network may be designed without regard to any reaction upon the cable.

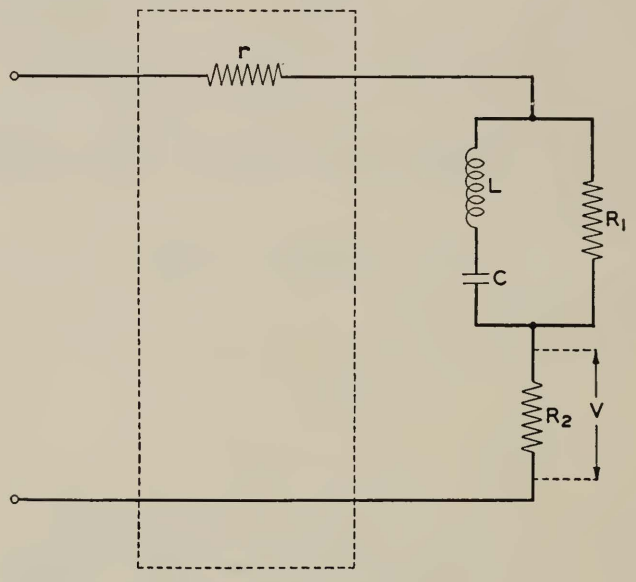


Fig. 8—Distortion corrective circuit for long telegraph cable

One section of the network of Fig. 8 with the values of the constants:

$$r + R_2 = 5{,}000 \text{ ohms}, \qquad L = 26 \text{ henrys}$$
$$R_1 = 51{,}100 \text{ ohms}, \qquad C = 1.56 \text{ microfarads}$$

is used when $l = 500$ miles. The phase of this network is shown in Fig. 5 also. Another equal network section may be added for each additional 500 miles of cable but there is no necessity, of course, for the sections to be equal. If it contains a one-way thermionic tube, each section may be added without affecting what has gone before, and the resultant phase angle will be simply the sum of all of the phase angles of the separate parts.

The improvement in the building-up of the indicial admittance accompanying the use of the phase compensator is evident from

Fig. 1 on comparing curve (2) with curve (1). Curve (2) is computed from the formula [3]

$$A(t) = \frac{2}{\pi}\int_0^\infty \frac{\alpha(\omega)}{\omega} \sin t\omega d\omega, \tag{17}$$

where $\alpha(\omega)$ is the real component of the transfer admittance of cable and network combined (equation (14)).

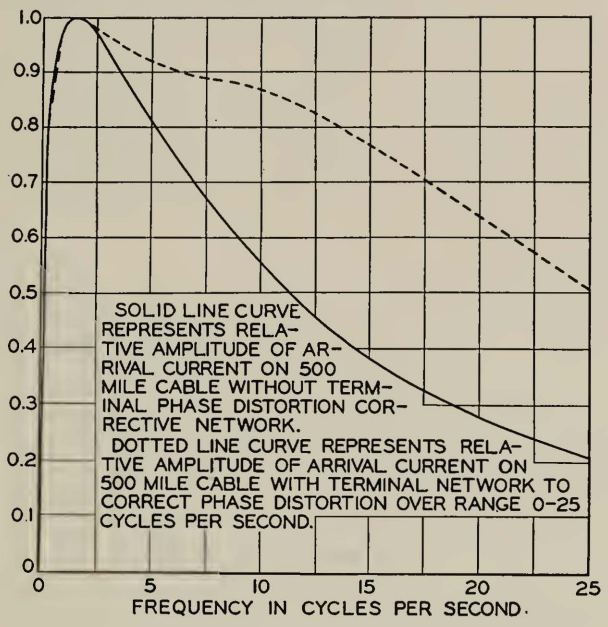


Fig. 9—Amplitude variation on long telegraph cable

The curves of Fig. 9 show that this network affords some attenuation equalization as well as very good phase correction. Amplitude and phase correction, as we have seen, are analytically independent processes. Nevertheless, some arrangements may, theoretically, be designed to correct amplitude and phase simultaneously. A method for so designing a network similar to the one under discussion at present has been developed by O. J. Zobel.[10] In such cases, however, in order to obtain physically desirable values in practical applications, it has usually been found necessary to design the network for one purpose, thereby automatically obtaining some improvement in the other respect, as in the present instance.

The maximum phase displacement obtainable with one section of

[10] This is discussed in a forthcoming paper by O. J. Zobel.

this network is $\pi/2$ radians. Thus, it becomes necessary to use more than one section on a long cable but it is also advantageous from the point of view of flexibility of design. By adopting a standard unit, for a 500-mile section of cable, for instance, the networks may be readily accommodated to different lengths of line.

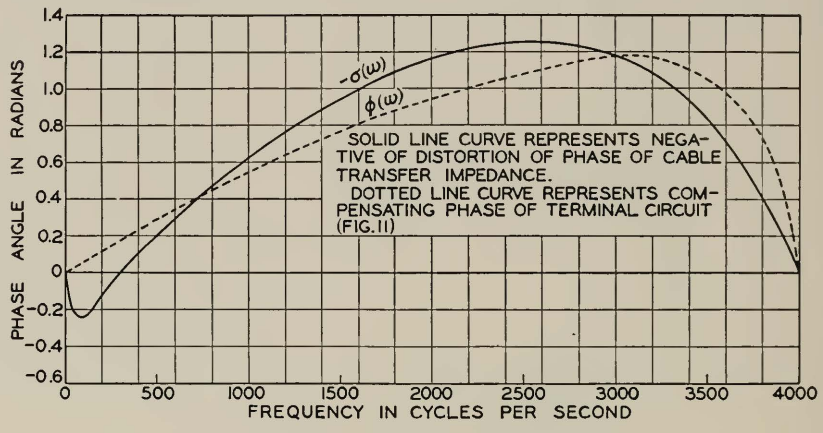


Fig. 10—Phase distortion correction on 20 mile 19 gauge H-44-25 loaded cable, $0 < f < 4000$

It is interesting to observe that the analytical expression

$$\phi(\omega) = \tan^{-1}\frac{ax}{1 + bx^2} \tag{15}$$

is positive when $0 < \omega < \omega_m$, and zero when $\omega = 0$ or ω_m, provided

$$x = \frac{\omega/\omega_m}{1 - \omega^2/\omega_m{}^2}. \tag{18}$$

Hence it will correct the phase distortion on the loaded cable as shown in Fig. 6. The required phase shift is obtainable in the type of network shown in Fig. 11. This network, however, has the disadvantage of tending to increase the attenuation distortion of the loaded cable rather than to equalize it.

When amplification is not required, it is undesirable to use the device of an amplifier in each section of network for the sole purpose of eliminating terminal reflection. As the characteristic impedance of a transmission line may usually be regarded as a constant resistance over the range of essential frequencies, the same purpose may be accomplished by applying networks whose characteristic impedance

is also a constant resistance of the same value as the characteristic impedance of the line. A number of general recurrent networks of the so-called 'constant resistance' type [11] are known and such a

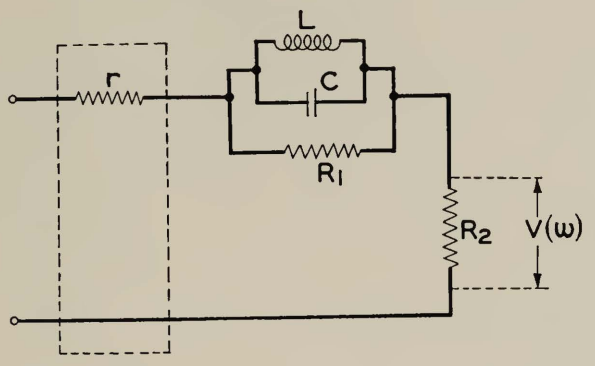


Fig. 11—Phase distortion corrective circuit for loaded cable. For 20 miles of 19 gauge H-44-25 loaded cable, $r + R_2 = 5000$ ohms, $R_1 = 101{,}000$ ohms, $L = 0.460$ henry, $C = 0.0216$ microfarad.

network [10] has been applied to correct the distortion on the submarine cable. The arrangement is shown in Fig. 12. It will be observed that z_{11} and z_{21} are inverse networks of impedance product R^2; that

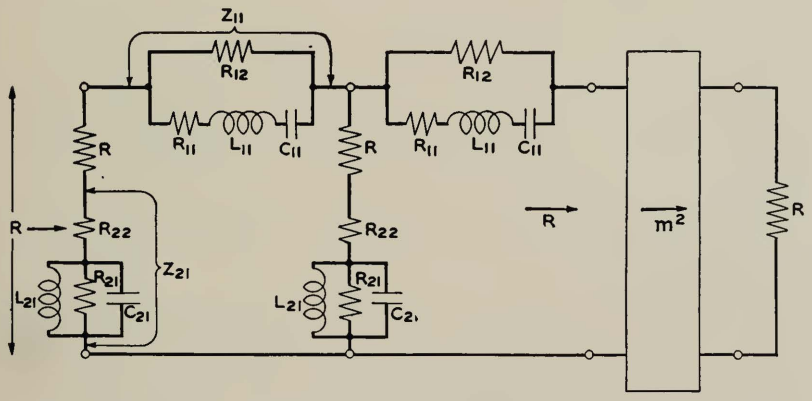


Fig. 12—Constant resistance type corrective circuit for long telegraph cable. m^2 represents distortionless amplifier

is, $z_{11}z_{21} = R^2$, which is, in general, the requisite relation among the network constants to provide a constant resistance characteristic impedance, R. In practical performance the networks of Figs. 8 and 12, each having the basic arrangement z_{11}, are equivalent. Thus the

[11] See reference 11.

function of eliminating terminal reflection, accomplished in the former by the vacuum tube, is accomplished in the latter by the additional impedance elements.

2. *Loading Systems*

The use of loaded cable circuits for the transmission of programs and for the transmission of pictures introduced some interesting problems in connection with phase correction at the higher frequencies. In this connection, a study of the building-up of sinusoidal oscillations in long loaded cables led to the establishment of two propositions which are of great value in comparing communication systems with respect to the quality of transmission.[12] These two propositions relate the variation with frequency of the steady state phase to the duration and nature of the transient distortion. They are applicable to all types of periodic loading, with or without terminal phase compensators under two restrictions; namely, (1) the line must comprise at least 100 loading sections and (2) the transducer as a whole must be approximately equalized as regards absolute steady state values of the received current in the neighborhood of the applied frequency. The successive derivatives of the total phase angle $B(\omega)$ with respect to ω will be denoted by $B'(\omega)$, $B''(\omega)$, $B'''(\omega)$. $B(\omega)$ may be understood to represent the sum of the phase differences due to transmission over the line and a terminal distortion corrective network, as well as the phase difference on the line alone. The propositions follow.

Case I: $B''(\omega) \neq 0$ and $\sqrt{B''(\omega)/2!}$ large compared with $\sqrt[3]{B'''(\omega)/3!}$.

The envelope of the oscillations in response to an e.m.f. $E \cos \omega t$ *applied at time* $t = 0$ *reaches 50 per cent. of its ultimate steady value at time* $t = B'(\omega)$ *and its rate of building-up is inversely proportional to* $\sqrt{B''(\omega)}$.

Case II: $B''(\omega) = 0$, $B'''(\omega) \neq 0$ and $\sqrt[3]{B'''(\omega)/3!}$ large compared with $\sqrt[4]{B^{IV}(\omega)/4!}$.

The envelope of the oscillations in response to an e.m.f. $E \cos \omega t$ *applied at time* $t = 0$ *reaches 1/3 of its ultimate steady value at time* $t = B'(\omega)$ *and its rate of building-up is inversely proportional to* $\sqrt[3]{B'''(\omega)}$.

These propositions furnish supplementary verification of the condition with regard to phase variation already established as a requirement for distortionless transmission; namely, that the phase vary linearly with the frequency or

$$B(\omega) = \omega\tau,$$

[12] See reference 4.

where τ is a constant. It follows immediately that

$$B'(\omega) = \tau.$$

Thus, when the steady state value of phase propagation is in linear relation with the frequency, all signals of any frequency whatever reach their proximate steady state in the same interval of time τ. These propositions make it possible to calculate two important criteria of the transmission properties of the line: (1) the variation with respect to frequency of the time interval τ required for the current to build up to its proximate steady state value and (2) its rate of building-up at time $t = \tau$. The first is very important in telephony but has no significance in telegraphy since only one frequency is transmitted, whereas the second is important in both telephony and telegraphy. While the formulas underlying these propositions are approximate, they are sufficiently accurate for a study of the comparative merit of different types of transmission systems or for the design of loaded lines.

For the purpose of reducing transient distortion on a long loaded cable of N sections, the loading and critical frequency $\omega_c/2\pi$ may be designed on the basis that the two time intervals,

$$t_1 = 2N/\omega_c \tag{19}$$

and

$$t_2 = \frac{2N}{\omega_c}\left(\frac{1}{\sqrt{1 - \omega^2/\omega_c^2}} - 1\right), \tag{20}$$

should be less than given quantities determined by experience. The two expressions t_1 and t_2 represent, respectively, (1) the time of transmission of d.c. current or the nominal time of transmission and (2) the excess of the time of building-up of the frequency $\omega/2\pi$ to approximately one half the steady state value over the nominal time of transmission.[13] The right hand member of formula (20), although initially determined by Carson by a method entirely dissimilar to that in his later investigation, is, it will be noted, given by

$$t_2 = B'(\omega) - B'(0) = T - T_0 \tag{21}$$

and

$$t_1 = B'(0) = T_0. \tag{22}$$

The nominal time of transmission t_1 is significant from the standpoint of echo effects but, with regard to phase distortion correction alone, emphasis is placed upon the quantity $t_2 = T - T_0$ as the more important factor. The enormous improvement in this respect of the

[13] See reference 2.

system of light loading over medium heavy loading is shown in Fig. 19. Below about 3,200 cycles, $T - T_0 < .0005$ second on a 50 mile light loaded line. It is obvious that the higher the frequency the greater the distortion.

Another loading system proposed as a means of providing desirable phase characteristics is the lattice loaded line.[14] This consists of the use of a section of the network of Fig. 14 as the periodic loading unit instead of the ordinary coil unit. Putting $C = rC_0$, where C_0 is the capacity of the line between loading units per section, we have for this system, when non-dissipative,

$$f_c = \frac{1}{\pi\sqrt{LC_0}}, \tag{23}$$

$$B(\omega) = 2N \sin^{-1} \frac{f}{f_c} \sqrt{\frac{1+r}{1+r(f/f_c)^2}}, \tag{24}$$

$$B'(\omega) = \frac{N}{\pi f_c} \sqrt{1+r} \frac{1}{[1+r(f/f_c)^2]\sqrt{1-(f/f_c)^2}} \tag{25}$$

and

$$T_0 = B'(0) = \frac{N}{\pi f_c} \sqrt{1+r}. \tag{26}$$

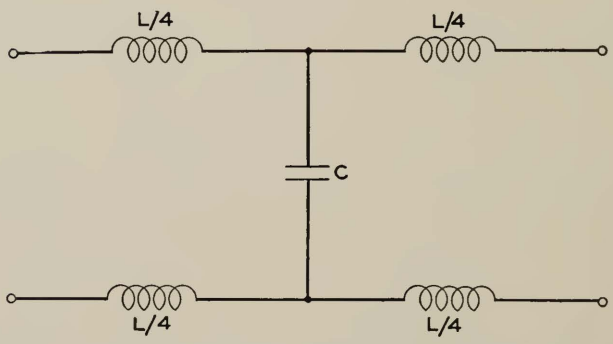


Fig. 13*a*—Section of standard non-dissipative coil loaded line. L = .044 henry, C = .0705 microfarad (for H-44 loading on 19 gauge side circuit.)

With the constants of the lattice loading system chosen to give approximately the same attenuation per mile as the standard loading, as in Figs. 13*a* and 13*b*, the time $t_2 = T - T_0$, it will be seen from Fig. 19, is less for it than that for the standard loading system for frequencies up to about 3,500 cycles but rapidly becomes greater thereafter. A combination of coil and lattice loading obtained by

[14] This was proposed by D. A. Quarles of the Bell Telephone Laboratories in unpublished memoranda.

alternating the coil and lattice loads affords some improvement in this respect. Another possible means of reducing the time interval $T - T_0$ at the higher frequencies is to reduce the spacing of the

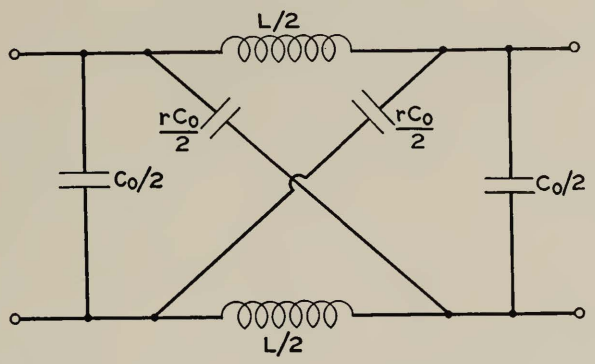


Fig. 13b—Section of non-dissipative lattice loaded line. L = .066 henry, C_0 = .0705 microfarad, r = .50 (19 gauge cable).

loading coils on the ordinary loaded line, keeping the attenuation per mile the same as before. As this change increases N and f_c in the same ratio, the effect is to reduce $T - T_0$.

3. *Lattice Type Terminal Network*

Instead of being used to replace the coil unit in the long periodically loaded line for the purpose of reducing the phase distortion, the lattice network [15] may be applied to the coil loaded line as a terminal phase compensator, a use to which it is peculiarly adapted by virtue of the nature of its characteristics.[16] These have been known for many years. The ideal non-dissipative simple lattice type recurrent network, shown in Fig. 14, with series inductance L and crossed capacity C per section, has a pure resistance characteristic impedance

$$K = \sqrt{L/C}, \tag{27}$$

a phase angle

$$\phi(\omega) = 2N \tan^{-1}(\omega\sqrt{LC}/2), \tag{28}$$

where N is the number of sections, and zero attenuation for all frequencies.[17] These relations readily follow from the general formulas (32)–(35) given below. The phase angle (Fig. 15), therefore, is

[15] The possible usefulness of the lattice network as a phase shifting device was pointed out in a general way by G. A. Campbell. See references 12 and 13.

[16] In this connection see references 12, 13 and 14 to the work of Karl Kupfmuller of the Siemens-Halske Company of Berlin, who independently applied the simple lattice network in the same way.

[17] See reference 13.

positive and varies from zero to π per section. Thus it is of a general form suitable to compensate for the distortion in the cable phase. One or more sections of the simple lattice network when properly

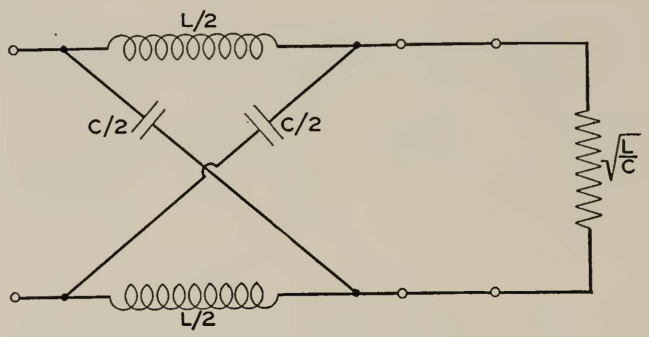


Fig. 14—Phase distortion corrective circuit for loaded cable: simple lattice type

terminated may then be connected directly to the loaded cable without appreciable reflection loss and without affecting the attenuation characteristic of the cable.

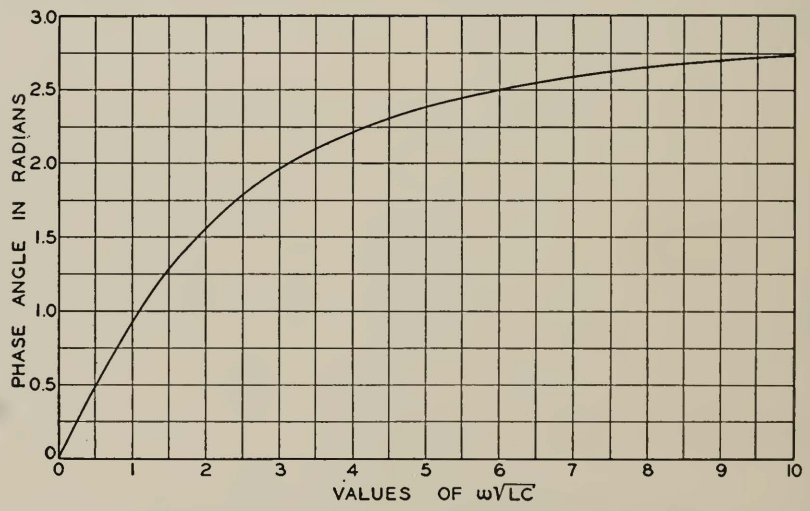


Fig. 15—Phase angle per section of simple lattice network

Proceeding to the transient aspect, the time T required for the current in response to the e.m.f. of frequency $\omega/2\pi$ to build up to 50 per cent. of its steady state value is given by

$$T = \phi'(\omega) = N\sqrt{LC}\,\frac{1}{1+\left(\frac{\sqrt{LC}}{2}\omega\right)^2} \tag{29}$$

and

$$T - T_0 = \phi'(\omega) - \phi'(0)$$

$$= N\sqrt{LC}\left(\frac{1}{1+\left(\frac{\sqrt{LC}}{2}\omega\right)^2} - 1\right). \tag{30}$$

The variation of $T - T_0$ in terms of the general parameter $\omega\sqrt{LC}$ is shown in Fig. 16. As T is a maximum at zero frequency and decreases with increasing frequency, its variation tends to equalize the delay on the cable. It will be demonstrated later that a more complicated type of lattice network is available to supplement the simple lattice type to improve the equalization still further.

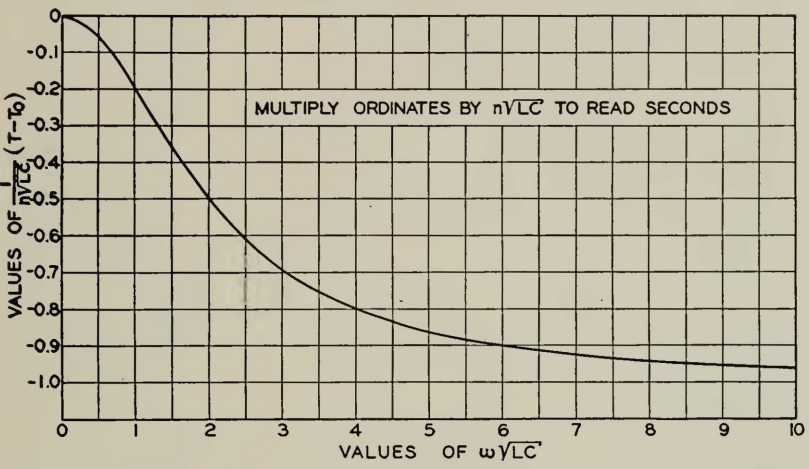


Fig. 16—Time of building-up, $T - T_0$, for simple lattice network of n sections

To provide partial delay equalization by means of the simple lattice network, phase angle variation alone need be considered without taking into account its derivatives. As pointed out above, the distortion $\sigma(\omega)$ of the cable phase $B(\omega)$ is given by

$$\sigma(\omega) = B(\omega) - \omega\tau \pm n\pi,$$

where τ is a constant representing the slope of the required distortionless phase. In the present case n will be zero. The constant τ also represents the time in which the current will build up to proximate steady state on the corrected cable. It is desirable, usually, that this time be as short as possible, and, moreover, the smaller the value of τ, the smaller will be the amount of distortion to be corrected,

remembering, of course, that $\sigma(\omega)$ is to be negative. On the other hand, a consideration of the nature of the phase variation, $\phi(\omega)$, of the lattice network as compared to the distortion of the cable phase

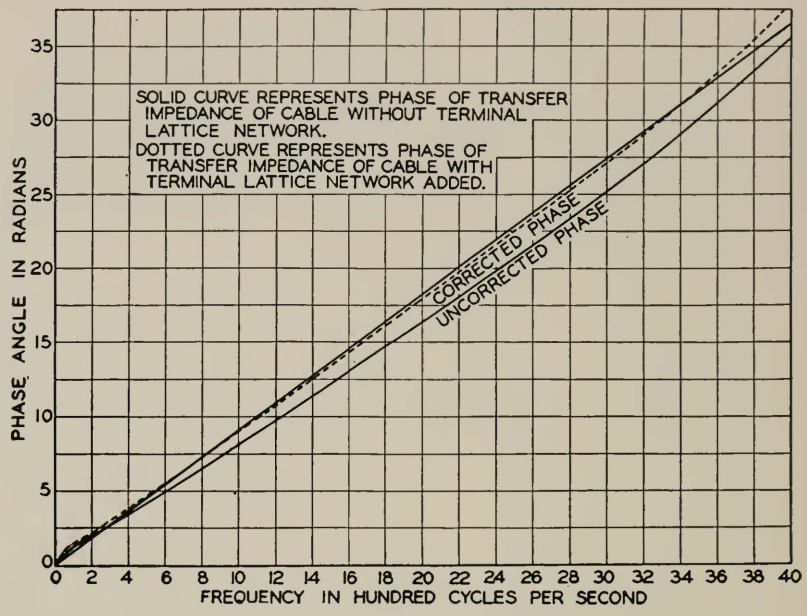


Fig. 17—Phase distortion correction on 25 miles of 19 gauge H-44-25 loaded cable

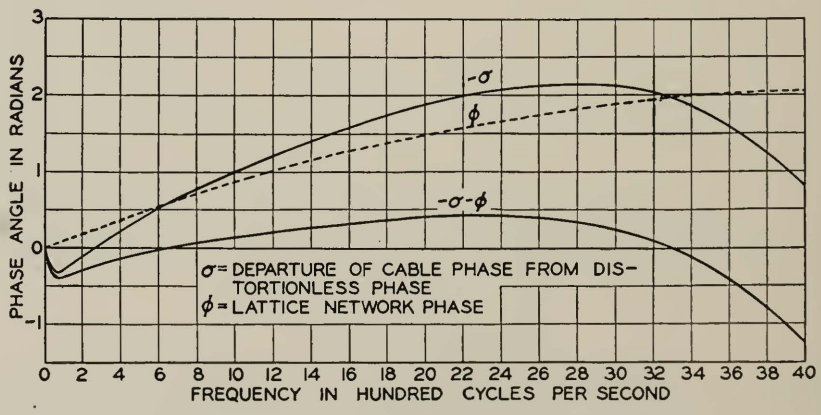


Fig. 18—Phase distortion correction on 25 mile 19-gauge H-44-25 cable

corresponding to various values of τ, leads to a choice of τ such that approximately the maximum cable distortion occurs at a frequency somewhat below the upper frequency of the essential range (Figs. 17 and 18).

While $\phi(\infty) = \pi$, it is apparent from Fig. 15 that ϕ increases very slowly for values of $\phi > 2.5$ and it is found that a distortion angle of not more than about 2.5 radians can well be compensated for with each section of lattice network. $\sigma_m/2.5$ determines roughly the number of sections required where σ_m represents the maximum distortion for the total length of line. It is only essential that the total number of sections be included where correction is desired. The corrective structure may be divided and one or more sections located at convenient points throughout the length of the cable whose phase is to be corrected. In the latter case a desirable arrangement on a repeatered cable is to insert at each repeater point a sufficient number of sections to correct the distortion on the cable length between two successive repeaters.

If the network is connected directly to the line, i.e., without the interposition of a unilateral element such as a vacuum tube, the necessary condition

$$K(\omega) = R,$$

where R is a constant representing approximately the characteristic impedance of the cable, imposes one limitation upon the constants L and C, leaving one other to be fixed by the phase. For this condition, put

$$\phi_a(\omega) = -\sigma_a,$$

where σ_a is the value of $\sigma(\omega)$ at a frequency $f_a = \omega_a/2\pi$ near the upper limiting frequency of the correction range and at which it is desirable to have exactly $\sigma(\omega) = -\phi(\omega)$. Substituting for K and ϕ_a in (27) and (28) and solving, gives

$$\begin{aligned} L &= -\frac{2R}{\omega_a}\tan\frac{1}{2}\sigma_a, \\ C &= -\frac{2}{\omega_a R}\tan\frac{1}{2}\sigma_a. \end{aligned} \tag{31}$$

An application of this method of design to the 19-gauge H-44-25 cable is shown in Figs. 17 and 18. This design requires two sections of lattice network at each repeater point of constants

$$L = .116 \text{ henry},$$
$$C = .181 \text{ microfarad}$$

per section, the distance between repeaters being 50 miles.

While the design above effects considerable improvement, it is

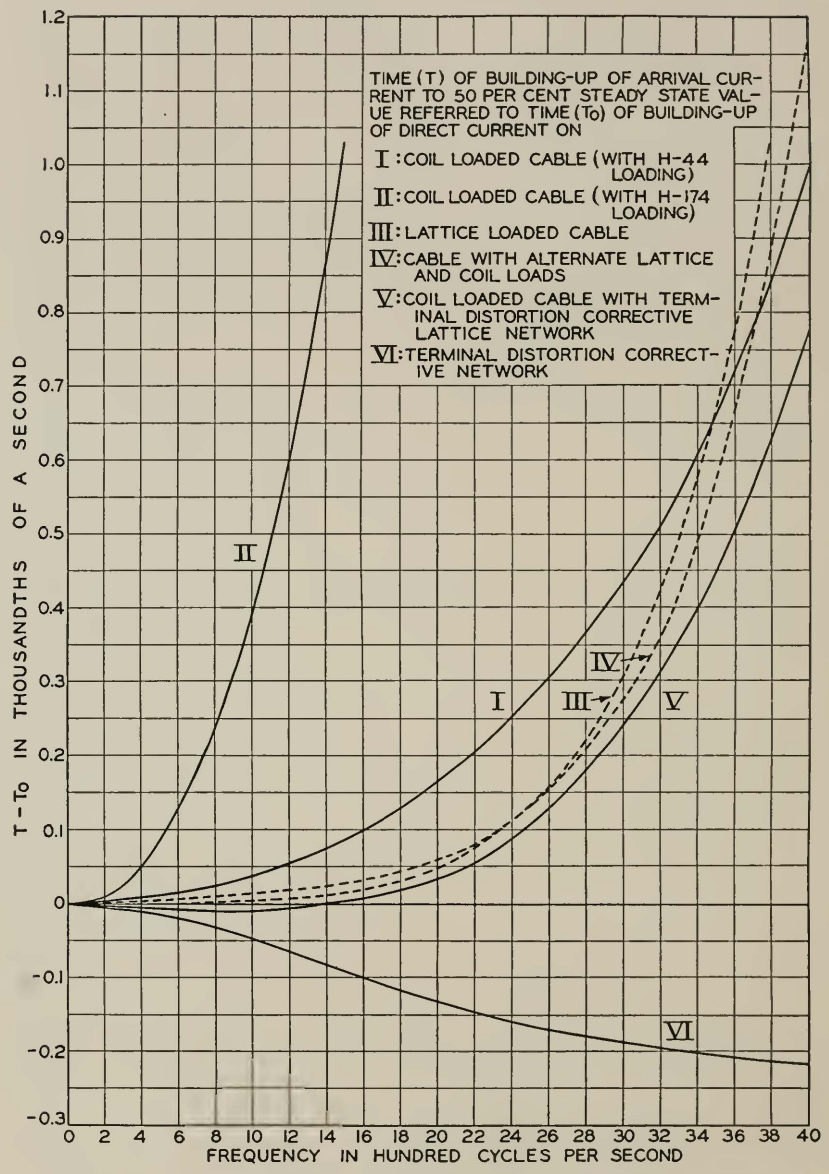


Fig. 19—Transient distortion on 50 miles of 19 gauge loaded cable

evident from a consideration of the duration of the transient distortion that the correction is not perfect, although it is appreciably better than that afforded by either lattice loading or a combination of lattice and coil loading. In Fig. 19 the delay in the time of building-up of any frequency $\omega/2\pi$ over the time of building-up of a direct current is shown for these different systems. Since the physical desideratum is a minimum constant delay for all frequencies, and the delay is quite sensitive to small deviations from the distortionless steady state phase angle, the former is probably a better basis of design and comparison than the latter.

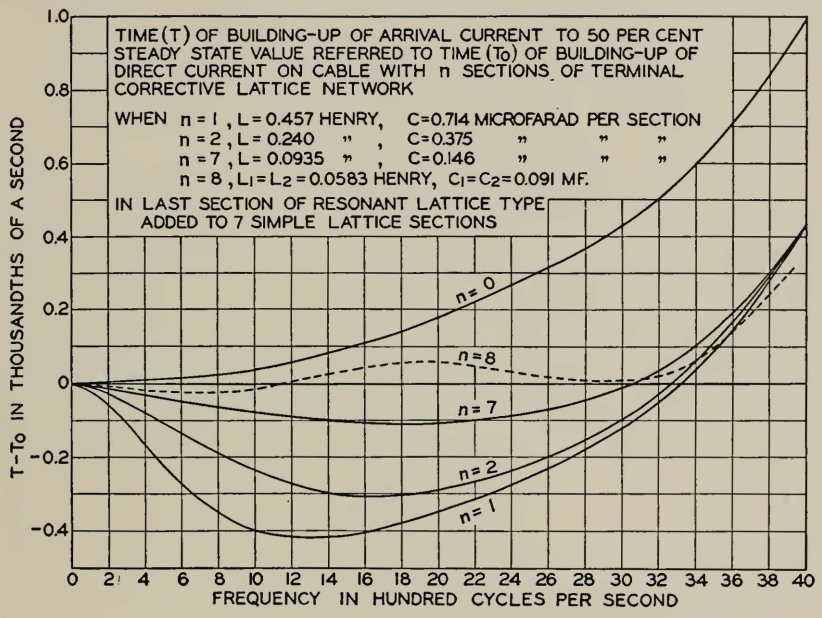


Fig. 20—Transient distortion on 50 miles of 19 gauge light loaded cable

Suppose that it is required to reduce the delay at 4,000 cycles to the value or to less than the value of the delay on the uncorrected cable at 3,000 cycles. The procedure will be to solve equation (30) for $\sqrt{LC}$ in terms of N and the required value of $T - T_0$ at the given frequency $\omega_a/2\pi = 4{,}000$, substitute these values in equation (30) and compute $T - T_0$ over the essential frequency range. It is immediately apparent from Fig. 20 that the improvement at the higher frequencies is at the expense of the intermediate frequencies where the delay is reduced too much. This disadvantage is lessened by increasing the number of sections but this method is uneconomical

and only partially successful because a saturation point is soon reached in the gain obtained with more apparatus.

This difficulty may be overcome by adding networks of the more complicated lattice type shown in Fig. 21.[18] This may appro-

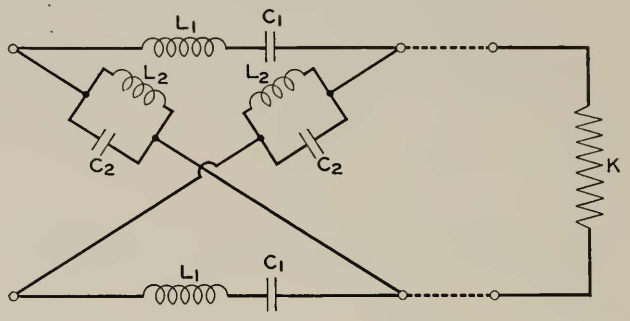


Fig. 21—"Resonant" lattice network

priately be called the 'resonant' lattice type. Its characteristics are most easily derived from the general lattice network having the impedance $z_1/2$ in each series branch and the impedance $2z_2$ in each diagonal shunt branch. Since the characteristic impedance, K, of any section of line is equal to the square root of the product of the open- and closed-circuit impedances and the propagation constant, Γ, per section, is equal to the anti-hyperbolic tangent of the square root of the quotient of the closed-circuit impedance divided by the open-circuit impedance,[19] we have, for the lattice network, in general,

$$\cosh \Gamma = 1 + \frac{2z_1}{4z_2 - z_1} \tag{32}$$

or

$$\tanh \frac{\Gamma}{2} = \frac{1}{2}\sqrt{z_1/z_2},$$

$$K = \sqrt{z_1 z_2}. \tag{33}$$

Thus the requirement that the characteristic impedance be a real constant will, in general, be fulfilled provided

$$z_2 = R^2/z_1, \tag{34}$$

where R is a real constant approximately equal to the characteristic impedance of the cable. This gives

$$\tanh \frac{\Gamma}{2} = \frac{z_1}{2R}. \tag{35}$$

[18] This suggestion was made by H. Nyquist.
[19] See reference 12.

Now if $z_1/2$ is the impedance of a series resonant circuit, we may write

$$\frac{z_1}{2R} = i\omega\frac{b}{2\omega_0} + \frac{b\omega_0}{i2\omega}, \tag{36}$$

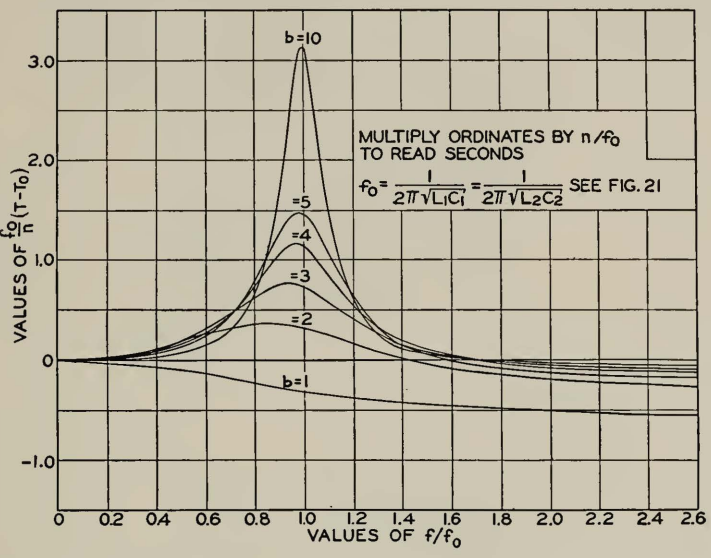


Fig. 22—Time of building-up, $T - T_0$, for resonant lattice type network of n sections

where b and ω_0 are constants. Then

$$\tanh\frac{\Gamma}{2} = \tanh i\frac{\phi}{2} = i\frac{b}{2}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right),$$

or

$$\phi = 2\tan^{-1}\frac{b}{2}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right), \tag{37}$$

$$T = \frac{\frac{b}{\omega_0}\left(1 + \frac{\omega_0^2}{\omega^2}\right)}{1 + \frac{b^2}{4}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)^2}, \tag{38}$$

and

$$T_0 = \frac{4}{b\omega_0}. \tag{39}$$

Then, from equations (34) and (36),

$$L_1 = \frac{bR}{2\omega_0}, \qquad C_1 = \frac{2}{b\omega_0 R}, \tag{40}$$

and

$$L_2 = \frac{2R}{b\omega_0}, \qquad C_2 = \frac{b}{2R\omega_0}.$$

The expression for T or $T - T_0$ will have a maximum at $\omega = \omega_0$ if b is large enough. The value of the maximum may be increased by increasing b and its location on the frequency scale may be changed by varying ω_0. A family of curves of $f_0(T - T_0)$ for different values of the parameter b are shown in Fig. 22 with the abscissa f/f_0.

To illustrate the combination of the resonant lattice type with the simple lattice type, add one section of the former with $b = 2$, $f_0 = 2{,}190$, to the seven sections of the latter already applied to the

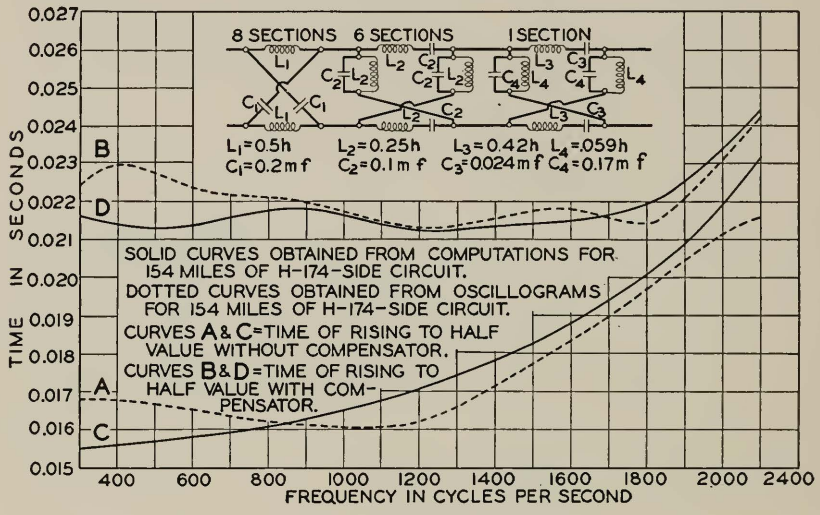


Fig. 23—Transients in loaded lines with and without phase compensator

cable (Fig. 20). The dotted curve for $T - T_0$, which results, is practically zero over most of the frequency range. The value of f_0 was determined so that the delay curve for the resonant lattice type would be complementary to the curve for the remainder of the circuit.

The combination of the two types of lattice network is effective in correcting even the large amount of phase distortion which results in transmission over the medium heavy loaded cable. Fig. 23 (which is reproduced by courtesy of Mr. H. Nyquist) shows a design for use on 154 miles of H-174 side circuit. The experimental results, obtained from oscillograms, are seen to be in close accord with the computed results.

The improvement due to correction of phase distortion by means of the phase compensator circuits employed for picture transmission is illustrated in Figs. 24*a*, 24*b* and 24*c*. Fig. 24*b* is from a negative which was transmitted over a 352-mile H-174 circuit without phase

correctors. Fig. 24*c* is the same print sent over the same circuit after the circuit had been equipped with phase correctors. Fig. 24*a*, which was sent locally, is shown for comparison. Fig. 24*c* is seen to be practically as clear as Fig. 24*a* and both are substantially better than Fig. 24*b*.

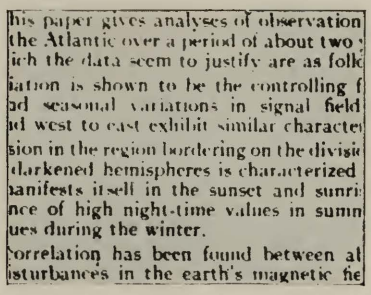

his paper gives analyses of observation
the Atlantic over a period of about two
ich the data seem to justify are as follo
iation is shown to be the controlling f
ıd seasonal variations in signal field
ıd west to east exhibit similar characte
sion in the region bordering on the divisi
darkened hemispheres is characterized
anifests itself in the sunset and sunri
nce of high night-time values in summ
ues during the winter.
Correlation has been found between al
isturbances in the earth's magnetic fie

Fig. 24*a*—Print transmitted locally

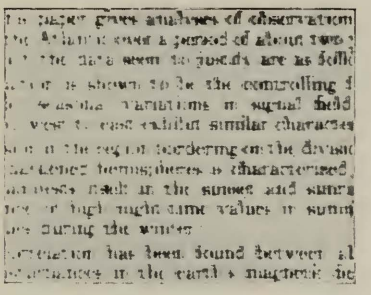


Fig. 24*b*—Print transmitted over 352 mile H-174 loaded cable without corrector

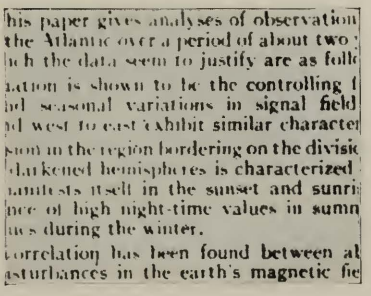

his paper gives analyses of observation
the Atlantic over a period of about two
ich the data seem to justify are as follo
iation is shown to be the controlling f
ıd seasonal variations in signal field
ıd west to east exhibit similar characte
sion in the region bordering on the divisi
darkened hemispheres is characterized
anifests itself in the sunset and sunri
nce of high night-time values in summ
ues during the winter.
Correlation has been found between al
isturbances in the earth's magnetic fie

Fig. 24*c*—Print transmitted over 352 mile H-174 loaded cable with terminal phase corrector

References

1. "Telephone Transmission over Long Cable Circuits." (A. B. Clark, *Jour. A. I. E. E.*, Vol. 42, No. 1, 1923, and *B. S. T. J.*, Jan., 1923.)
2. "Loading System." (Carson, Clark and Mills, U. S. Patent No. 1,564,201, Dec. 8, 1925.)
3. "Theory of the Transient Oscillations of Electrical Networks and Transmission Systems." (Carson, *Trans. A. I. E. E.*, Feb., 1919.)
4. "Building-up of Sinusoidal Currents in Long Periodically Loaded Lines." (Carson, *B. S. T. J.*, Oct., 1924.)
5. "Distortion-Correcting Circuit." (Carson, U. S. Patent No. 1,315,539, Sept. 9, 1919.)
6. "Receiving System for Telegraphic Signals." (Mathes, U. S. Patent No. 1,586,821, June 1, 1926.)
7. "The Loaded Submarine Telegraph Cable." (Buckley, *B. S. T. J.*, July, 1925.)
8. "The Application of Vacuum Tube Amplifiers to Submarine Telegraph Cables." (Curtis, *B. S. T. J.*, July, 1927.)
9. "Electric Circuit Theory and the Operational Calculus." (Carson, McGraw-Hill Book Co., 1926, 1st ed., p. 185.)
10. "Attenuation Equalizer." (Hoyt, U. S. Patent No. 1,453,980, May 1, 1923.)
11. "Electrical Network and Method of Transmitting Electrical Currents." (Zobel, U. S. Patent No. 1,603,305, Oct. 19, 1926.)
12. "Physical Theory of the Electric Wave-Filter." (Campbell, *B. S. T. J.*, November, 1922.)
13. "Maximum Output Networks for Telephone Substation and Repeater Circuits." (Campbell and Foster, *Trans. A. I. E. E.*, Vol. 39, 1920, p. 258.)
14. "Die Technik der Telegraphie und Telephonie in Weltverkehr." (Lüschen, *E. T. Z.*, July 31, 1924.)
15. "Über Einschwing Vergänge in Pupinleitungen und ihre Verminderung." (Küpfmüller und Mayer, Wissenschaftlich Veröffentlichungen aus dem Siemens-Konzern, Vol. V, Sect. 1, 1926.)
16. "Die Erhöhung der Reichweite von Pupinleitungen durch Echosperrung und Phasenausgleich." (Küpfmüller, *E. N. T.*, Vol. 3, No. 3, 1926.)

High-Speed Ocean Cable Telegraphy

By OLIVER E. BUCKLEY

Synopsis: The invention of permalloy and its application to submarine cables have led to the installation of transoceanic cables of many times the traffic-carrying capacity of the former non-loaded cables. This paper relates briefly the history of the development of permalloy-loaded cables and discusses certain outstanding problems concerned with their design, construction and operation. In a concluding general survey the field of usefulness of loaded submarine telegraph cables is considered.

To a considerable extent the paper is a critical summary of material previously published by members of the staff of the Bell Telephone Laboratories. Its scope is indicated by the sub-titles as follows:

Loaded Cables Now in Service
Historical Remarks
Permalloy and Its Application to Cables
Principles of Design of Loaded Cables
Principles Involved in Operation
Apparatus for Restoration of Signals
Apparatus for Automatic Operation
Electrical Measurements of Loaded Cables
A General Survey

VOLTA devised his famous pile in 1799. Less than 60 years later, in 1858, the first telegraph message was sent over an Atlantic cable. Now nearly 70 years have passed since the remarkable feat of transatlantic telegraphy was first accomplished. Although the art of cable telegraphy may therefore be considered old, it cannot be said ever to have stopped growing. At all periods of its growth it has offered an interesting field for technical endeavor. An added interest was attached to it a little more than 25 years ago when Marconi, by his famous demonstration of transatlantic radio telegraphy, introduced a competitor. With the birth of this new child of science arose the question as to whether the art of telegraphing over cables would not ultimately die. But radio too required time to grow, and it is only within very recent years that there has been occasion for serious concern as to the future of the older art. Now a new advance has been made on the side of the cables and the race for supremacy in transoceanic communication has taken a new turn. The advance to which I refer is the introduction of the high-speed permalloy-loaded cable, and it is with regard to this advance that I wish to speak.

My object is to tell briefly what has been accomplished with cables of the permalloy-loaded type and to describe some of the outstanding features of development which have led to this accomplishment. No attempt will be made in what follows to discuss all phases of cable design and construction, but my remarks will be confined principally to those aspects of cable telegraphy with which the work in the Bell Telephone Laboratories has been concerned.

Loaded Cables Now in Service

There are at present seven high-speed ocean cables of the permalloy-loaded type in operation. Together they have a length of nearly 15,000 miles, which represents about five per cent of the total ocean cable mileage of the world. Their location and lengths are shown on the map in Fig. 1. With the exception of the Cocos Island-Perth (Australia) cable of the Eastern Extension Telegraph Company, these loaded cables are comprised in three transoceanic lines, two crossing the Atlantic and one crossing the Pacific.

The first loaded ocean telegraph cable was the New York-Horta (Azores) cable of the Western Union Telegraph Company which was laid in September 1924. The great success attained with it led to the installation of others, among which was the 1926 Horta-Emden cable of the Deutsch Atlantische Telegraphengesellschaft. The New York-Horta-Emden line thus formed provides not only for carrying a large volume of messages between America and Germany, but also gives a connection with the Italian cable at Horta. This line is now provided with a 5-channel multiplex printing telegraph equipment which is operated at a speed of about 1500 letters per minute and is expected ultimately to be operated at a considerably higher speed. Four of the five channels of this line provide direct communication between New York and Emden, and the fifth serves for messages relayed at Horta.

A second transatlantic line is formed by the New York-Bay Roberts and Bay Roberts-Penzance cables of the Western Union Telegraph Company which were laid in 1926. Each of these cables is capable of carrying more than 2500 letters per minute, but at present this line is being operated at only about one half that speed. The construction of operating equipment to realize the full 2500 letters per minute is now in process. The combined traffic-carrying capacity of these two transatlantic loaded lines is nearly as great as that which was previously provided by the sixteen older non-loaded cables which served to connect North America with Europe prior to 1924.

The Pacific Cable Board, also in 1926, installed loaded cables to parallel its non-loaded cables of 1902, connecting Bamfield, Fanning Island and Suva. A speed of over 1200 letters per minute has been reported for each section of this transpacific line. This speed is nearly four times that which was afforded by the older non-loaded cables over the same route.

Such an extension of facilities for transoceanic communication would be expected to have a pronounced effect on both the cost of

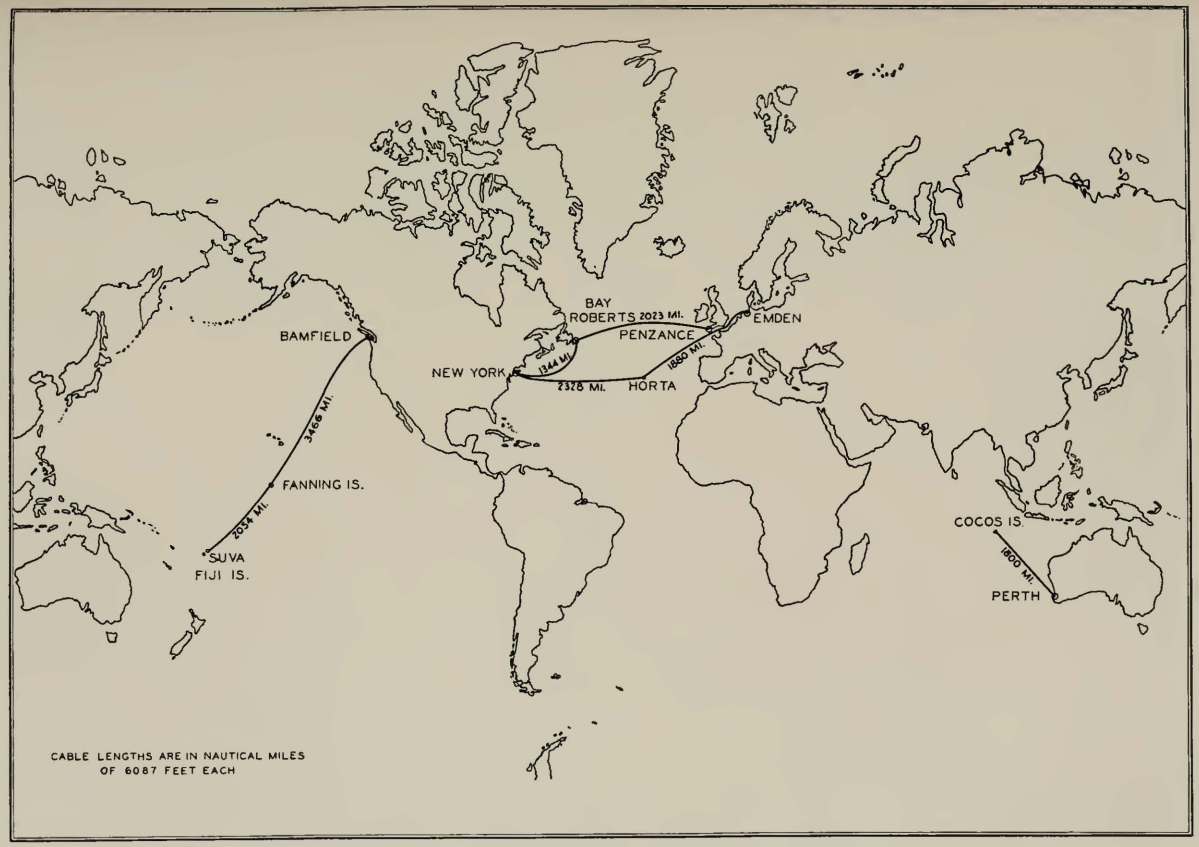


Fig. 1—Map showing location of loaded ocean cables

communication and the amount of communication between continents and, indeed, such an effect has already been experienced, significant reductions in rates having been made within the past year. In view of what has already been accomplished it seems not unlikely that the further introduction of loaded cables will completely revolutionize the whole transoceanic communication situation.

Historical

Like nearly all other important technical advances, that of the loaded telegraph cable is the outgrowth of contributions of many investigators and engineers working in different fields of endeavor. The history of the development of the idea of inductively loading transmission lines, from its original conception to recent times, is well known and need not be gone into here. An excellent theoretical analysis of the problem of transmitting signalling impulses over an ocean cable has been made by Malcolm, who in 1917, in his book on "The Theory of Submarine Telegraph and Telephone Cables," went so far as to predict that heavy continuous loading was the next great advance in the telegraph-cable art to be expected.

The practical accomplishment of the permalloy-loaded cable came about as a result of researches conducted in the laboratories of the American Telephone and Telegraph and Western Electric Companies, now known as the Bell Telephone Laboratories. Our interest in the problems of submarine cables was a natural part of our interest in all phases of electrical communication and was at first concerned principally with the application of vacuum tube amplifiers to ordinary telegraph cables. Considerable progress in the development of amplifiers for this purpose was made in the laboratory as early as 1914. The great demand for cable communication which the World War brought about led to further activity in this direction and to extensive tests of vacuum tube amplifiers on the cables of the Western Union Telegraph Company.

From these tests it was found that although the vacuum tubes would provide any desired amplification of signal strength and although by the combination of vacuum tubes and suitable electrical networks the distortion of the received signals could be corrected to any desired degree, relatively little actual gain in traffic capacity could be obtained by these means, since the real limit to cable speed was not distortion but interference. Although some improvement in cable speed could have been achieved by refinement of means for duplex operation and by improved means to eliminate extraneous interference, it was quite apparent that to obtain any great advance over the existing art would require a modification of the cable itself.

For over fifty years the cable had remained substantially unchanged in character, though great advances had been made in methods of operation. Inductive loading, which was proposed by Heaviside in 1887, was the obvious means for obtaining increased cable speeds, but no one had found a way to realize the advantages of loading as applied to ocean cables. Loading with evenly spaced coils as proposed by Pupin and used on land lines presented difficulties in laying and maintenance which practically prohibited this method. Continuous or Krarup loading by a wrapping of iron wire around the conductor of the cable was mechanically feasible but the amount of inductance which could be obtained in this way was not sufficient to justify its use. In order to make continuous loading advantageous for long ocean cables there was needed a material which could be much more easily magnetized than iron. Fortunately we had at hand as an aid to solving this problem the extraordinary magnetic material, permalloy, an alloy possessing magnetic permeability many times that of iron, even at the low magnetizing forces produced by the feeble currents of a telegraph cable. It is on permalloy that the loaded cable depends primarily for its success.

Although permalloy provided the means to give the cable the desired high inductance, the mere wrapping of this metal around the copper conductor of the cable was far from providing a practical solution of the problem of high-speed ocean telegraphy. To achieve this solution required the solution of very many subsidiary problems, concerned not only with the making of a cable but also with the transmission of signals over it and with the means for its practical operation. Work on all these phases of the problem of the loaded cable was actively pursued in our laboratories and in the field from the time of the first proposal, made in July 1919, to load a trans-oceanic cable with permalloy to the successful completion and operation of the New York-Horta cable.

During the first two years of this period our investigations were conducted wholly in the laboratory where hundreds of experimental lengths of loaded conductors were made and tested to convince ourselves that a permalloy-loaded cable could be manufactured and laid successfully. During the same period studies of the signal distortion of a loaded artificial cable and means for correcting distortion were carried on. Simultaneously methods of high-speed operation of loaded cables were developed, with the result that we were convinced from our laboratory experiments, not only that a permalloy-loaded cable could be made and laid successfully, but that it could also be operated commercially at the high speeds which we had predicted.

Having gone this far in the laboratory, it was decided to bring the results of our investigations to the attention of one of the cable operating companies with the object of securing a practical trial of a permalloy-loaded ocean cable.

On being shown what could be accomplished with permalloy loading, the Western Union Telegraph Company was quick to take advantage of this means of extending its cable facilities, and shortly thereafter arrangements were made whereby the Telegraph, Construction & Maintenance Company, Ltd., was to manufacture a cable for the Western Union Telegraph Company, using permalloy loading material supplied by the Western Electric Company and applied and treated under the direction of Western Electric engineers.

As a part of this undertaking, it was decided that prior to laying a complete transoceanic length of cable it would be desirable to make, lay and test a shorter length in order to obtain experience in manufacture and a test of its mechanical and electrical properties after it had suffered the extreme treatment to which a deep-sea cable is subject in laying. Accordingly, for such an experiment, 120 miles of cable of the same type which it was proposed to use for a transoceanic length was laid in a loop from the south shore of Bermuda in October 1923. Very thorough tests were made jointly by Western Electric and Western Union engineers to determine whether the electrical characteristics had been affected by laying and what attenuation and distortion were actually produced by such a cable. The results obtained were in excellent agreement with our predictions, and accordingly manufacture of the full 2300 miles required to connect New York and Horta was at once undertaken.

The New York-Horta cable which, like the 120-mile trial cable, was manufactured by the Telegraph, Construction & Maintenance Company with permalloy loading material applied and treated under the technical direction of Western Electric engineers, was laid in September 1924. Within an hour after the cable had been turned over to our engineers for test, a speed of 1500 letters per minute was obtained with the terminal apparatus which had been designed and provided in advance. In this case the messages were received on a high-speed siphon recorder of special design. Shortly thereafter, with the same apparatus, a speed of over 1900 letters per minute was obtained.

The speed of the cable having been demonstrated, commercial operation was quickly established with temporary operating equipment utilizing siphon recorders in conjunction with vacuum tube amplifiers. This type of operation was continued for about two

years during which the engineers of the Western Union Company and the Bell Laboratories worked together on the development of a multi-channel printing telegraph system which would adapt the operating methods previously developed by the laboratories to the needs of the telegraph company. In October 1926 the present five-channel printing telegraph apparatus was put into use.

Demands for other high-speed loaded cables quickly followed the successful demonstration of the New York-Horta cable. The Western Union Company arranged for the manufacture of the cables for the New York-Bay Roberts-Penzance route by the Telegraph, Construction & Maintenance Company, Ltd. On these cables permalloy supplied by the Western Electric Company was again used. The Norddeutsche Seekabelwerke A-G. arranged with the Western Electric Company for the supply of permalloy loading material and for technical assistance to manufacture the Horta-Emden cable for the Deutsch Atlantische Telegraphengesellschaft. The Pacific Cable Board arranged to have the cables for its Bamfield-Fanning Island-Suva line manufactured by two British companies and obtained licence therefor from the Western Electric Company. The shorter section from Fanning Island to Suva was made by Siemens Brothers & Co., Ltd., with permalloy supplied by the Western Electric Company and applied and treated with the technical direction of their engineers. The long section from Bamfield to Fanning Island was made by the Telegraph Construction & Maintenance Company, Ltd. All of these cables were laid in 1926.

Permalloy and Its Application to Cables

Permalloy, which made possible this radical change in the cable art, is the invention of G. W. Elmen of the Bell Telephone Laboratories. Since descriptions and explanations of some of its properties have already been given in papers by various members of the Bell Laboratories' staff,[1] the present discussion will be limited to a brief statement of its outstanding characteristics which are of consequence in connection with its use on cables.

The name "permalloy" has been applied to alloys of iron and nickel of more than about 30 per cent nickel content characterized by extraordinarily high magnetic permeability at very low magnetizing

[1] H. D. Arnold and G. W. Elmen, *Jour. Franklin Inst.*, Vol. 195, pp. 621–632, May 1923; *B. S. T. J.*, Vol. II, No. 3, p. 101.
O. E. Buckley and L. W. McKeehan, *Phys. Rev.*, Vol. 26, pp. 261–273, Aug. 1925.
L. W. McKeehan, *Phys. Rev.*, Vol. 26, pp. 274–279, Aug. 1925.
L. W. McKeehan and P. P. Cioffi, *Phys. Rev.*, Vol. 28, pp. 146–157, July 1926.
L. W. McKeehan, *Phys. Rev.*, Vol. 28, pp. 158–166, July 1926.

forces. The manner in which the initial permeability of these alloys varies with composition, when heat-treated in a particular way, is shown in Fig. 2, which is taken from the Arnold and Elmen paper. The magnetic properties of the alloys of this series depend in an extraordinary degree on their previous mechanical and thermal history. In general, high initial permeability is obtained by rapid cooling after a thorough softening of the metal by heating at a high temperature,

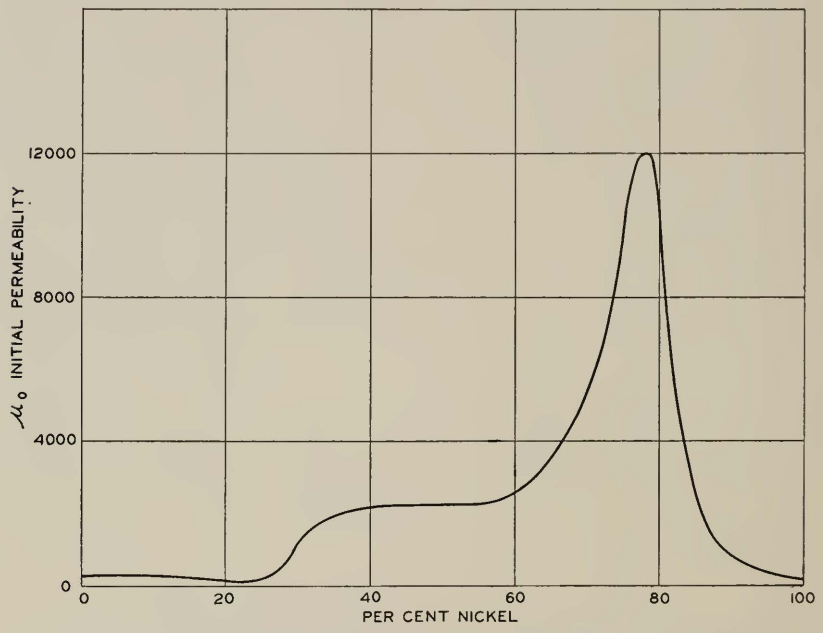


Fig. 2—Variation of initial permeability with composition of permalloy

this effect of rapid cooling being particularly marked on the compositions in the region of 80 per cent nickel. By control of the composition and heat treatment an initial permeability of more than 12,000 has been obtained with an alloy of $78\frac{1}{2}$ per cent nickel and $21\frac{1}{2}$ per cent iron, whereas iron or nickel alone ordinarily have initial permeabilities of only about 200 or 300. It is the high initial permeability of permalloy that is most important in its use on cables, though such an initial permeability as 12,000 would be even higher than is generally desired for a telegraph cable. For use on cable conductors permeabilities of the order of from 2000 to 5000 have been desired and obtained in practice.

Another important property of permalloy with regard to its use on cables is its resistivity, since high resistivity prevents excessive eddy-

current loss. The resistivity of the whole nickel-iron series of alloys is higher than that of either iron or nickel. By adding a third element, for example chromium, to the nickel and iron and keeping the ratio of nickel to iron about 4 : 1, a combination of very high resistivity and very high initial permeability may be obtained in the same alloy.

The permalloy used in the New York-Horta cable contained about 79 per cent nickel and 21 per cent iron with a small amount of manganese to make it more malleable. The permeability of this alloy as used on the New York-Horta cable was about 2300, its resistivity being about 16 microhm-cms. On the Horta-Emden cable, the New York-Bay Roberts-Penzance cables and the Fanning Island-Suva cable permalloy containing about 80 per cent nickel, 17.5 per cent iron, 2 per cent chromium and 0.5 per cent manganese was used. With this alloy an initial permeability of about 3700 was obtained. Its resistivity is about 38 microhm-cms.

The permalloy loading material used on the New York-Horta cable was in the form of a thin tape 0.006 inch (0.015 cm.) thick and 0.125 inch (0.32 cm.) wide applied in a closely wound helix surrounding the conductor. On the Horta-Emden cable tape 0.0059 × 0.098 inch (0.015 × 0.25 cm.) was used. On the New York-Bay Roberts and Bay Roberts-Penzance cables the tape thickness was 0.0055 inch (0.014 cm.) and the widths were 0.079 inch (0.20 cm.) and 0.123 inch (0.31 cm.) respectively. On the Fanning Island-Suva cable the permalloy is in the form of a wire of 0.011 inch (0.028 cm.) diameter applied in a single closely laid helix. The northern section of the Pacific cable from Bamfield to Fanning Island is reported [2] to be loaded similarly with "Mumetal" wire of 0.010 inch diameter, made by the Telegraph, Construction & Maintenance Company.

Very good results have been obtained with both tape and wire loading. The tape has the advantages of costing less to apply and of possessing greater mechanical strength, whereas the wire has the advantages of lower eddy-current loss and of being less affected by the earth's magnetic field. With either wire or tape loading the component of the earth's magnetic field parallel to the cable sets up magnetic induction in the helical loading material and consequently reduces its effective permeability for the small magnetizing forces of the signalling current. This reduction of effective permeability by the earth's magnetic field is greater, the greater the angle of lay with which the loading material is applied. Consequently this effect is generally greater with tape loading than with wire loading. Whether

[2] E. S. Heurtley, *Electrician*, Vol. 98, pp. 348–350, Apr. 1, 1927. See also *P. O. Elec. Engrs. Jl.*, Vol. 20, pp. 36–40, Apr. 1927.

tape or wire should be used is, in the end, an economic problem since any disadvantage of one with regard to the other may be compensated for by increasing the size of the copper conductor.

Permalloy has another property which it is important to consider in connection with its use on cables, namely, its great sensitiveness to mechanical strain. Strain of deformation applied to it will modify its magnetic characteristics, and very great changes in its permeability for small magnetizing forces may be produced by strains well within the mechanical elastic limit. Consequently in making the cable it is necessary to insure that the permalloy shall be as free as possible from strains of deformation. There are two principal ways in which the permalloy used for loading may be subject to such strains. The first comes in the manufacture of the loaded conductor and the second in the laying of the cable.

Since permalloy is so strain-sensitive it must be annealed after it has been applied to the conductor. Accordingly the hard-worked metal is wrapped around the copper conductor and the conductor is thereafter passed continuously through a furnace, maintained at approximately 900° C., and from the furnace into a cooling tube. The lengths of the furnace and cooling tube and the rate of passage of the conductor are so chosen as to insure that the loading material will get the necessary softening in the furnace and will be cooled at the proper rate in the cooling tube. Even though the permalloy is thus annealed on the conductor, it still might well be subject to considerable strain, since the copper, on being heated to such a high temperature, expands more than the permalloy and tends to weld to it and, on contracting, would bend the permalloy tape near the spots where welding occurs. To prevent this action the loading material is applied very loosely and means are taken to prevent adhesion of the permalloy to the copper. In spite of the great sensitiveness of the permalloy to mechanical strain, the loaded conductor after heat treatment stands ordinary handling very well without much loss of permeability. However, if it were insulated by the methods which have been used in the past in making deep-sea cables, it would lose much of its inductance on laying on account of the effect of the great pressures to which a cable is subjected.

To prevent reduction of the permeability and consequent loss of inductance on laying, it is necessary to provide that pressure on the insulating material shall produce only true hydrostatic pressure on the permalloy with no tendency to deform it. This result has been accomplished by vacuum-impregnating the permalloy-loaded conductor with a semi-fluid compound which fills all the interstices of

the conductor and also forms a layer a few thousandths of an inch thick on the outside of the loading material. The gutta-percha insulation may then be extruded over the impregnated conductor with the assurance that the semi-fluid compound will serve to equalize the pressure on the permalloy. Numerous compounds have been proposed and used for this purpose, that on the New York-Horta cable being of an asphaltic type. It is essential, of course, that this compound be sufficiently viscous at temperatures at which the gutta-percha is applied to permit extruding the gutta-percha around it and that it will also be sufficiently fluid at the temperature of the sea bottom, which may be as low as 2° C., to permit readjustment of the pressure on the permalloy. When a loaded conductor insulated in this manner is subjected to high pressures at low temperatures, it

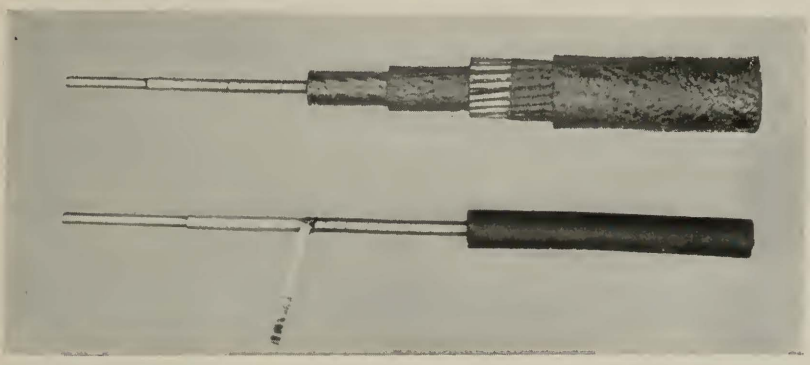


Fig. 3—Permalloy-loaded cable. Above, section of deep-sea type from New York-Horta cable. Below, section of core showing permalloy tape partly unwound

may be found that the inductance drops when the pressure is first applied but in a few minutes the compound flows so as to equalize the pressure on the permalloy and the inductance quickly comes back to its original value.

Outside of the insulated conductor or "core" of the cable the permalloy-loaded cables which have been made are quite like ordinary non-loaded cables, so there is no need to go into the further details of cable construction here. Fig. 3 shows a section of the deep-sea portion of the New York-Horta cable.

Principles of Design of Loaded Cables

There are two principal aspects of the design of a submarine cable—mechanical and electrical. Mechanically, the cable must be so designed as to insure that the conductor shall be continuous, that its

insulation shall be maintained and that the core, which comprises the conductor and insulation, shall not be damaged in laying or in subsequent repairing operations. Electrically, the cable must be so designed that it will serve properly to transmit signals. Thus the electrical design is concerned with the size of the conductor and the amount and characteristics of the loading material and insulation, whereas the mechanical design is concerned with the mechanical characteristics of the conductor and its insulation and with the jute and armor wire which serve to protect the core and give the cable the necessary strength. These two aspects of design cannot, of course, be considered quite independently of each other and both are in the ultimate analysis controlled by economic considerations. It is convenient for our present purposes, however, to consider them separately.

The mechanical design of cables is a well-established art and so great are the difficulties of laying and maintaining cables, even under the most favorable conditions, that it is desirable to avoid taking any liberties with this phase of cable construction. Fortunately the method of loading by a continuous wrapping of magnetic tape or wire introduces no need for radical change in the important mechanical features of the cable.

The copper conductor of the loaded cable is, as in many non-loaded cables, composed of a central copper wire surrounded by several flat copper strips. This form of conductor is flexible and economical of space, and the fact that it has several strands reduces the chance of a complete break. The loading tape or wire furnishes additional protection in this regard.

The thickness of gutta-percha must be sufficient to insure the integrity of the insulation at all points. It is, in fact, this consideration which established the amount of insulation used on the loaded cables which have been laid, since consideration of the theoretical economic optimum thickness of gutta-percha would in each case have demanded less gutta-percha than is considered safe. In this regard the insulating problem of the loaded cable is like that of the non-loaded cable.

The disposition of jute and armor wire around the core is determined wholly by mechanical considerations as in the case of the non-loaded cable for which the practice is fairly well standardized. Unlike the non-loaded cable, however, the loaded cable, in its electrical behavior, is affected somewhat by the presence and character of the armor wire as will be described later.

As is well known, the electrical behavior of *non-loaded* cables is determined almost wholly by their resistance and capacity and consequently the only important features to consider from the electrical

standpoint have been the size of the conductor and the thickness of insulating material. The electrical design of a loaded cable is, however, somewhat more complicated since in addition to copper resistance and electrostatic capacity we have here to be concerned with the inductance added by the loading material and also with added resistance factors which are introduced by its use. The problem of electrical design, therefore, involves determining not only the size of the copper conductor, but also the electrical and magnetic characteristics and the shape and dimensions of the loading material, as well as the electrical characteristics of the insulating material, which will give the highest speed of operation consistent with the mechanical and cost limitations which are imposed.

Since the object of the electrical design is to secure high operating speed, it is essential to consider what are the factors which limit speed and how they are taken into account. This subject has already been treated in some detail in previous papers,[3] and only a general review of the principal factors involved in the electrical design will be undertaken in the present paper.

In the history of cable development prior to the introduction of the permalloy-loaded cable various physical factors at different times limited the speed of operation which could be obtained with a long ocean cable. These were principally distortion of signals, sensitivity of receiving apparatus, limited safe sending voltage, inaccuracy of duplex balance, and extraneous interference from both natural and man-made sources. With the development of cable amplifiers and of improved means of signal shaping, the factors of distortion and limited sensitivity of receiving apparatus were effectually eliminated and at the time when the development of the permalloy-loaded cable was undertaken the speed of long cables was limited in most cases by the accuracy with which artificial lines could be made to balance cables in duplex operation. In some cases where extraneous interference was unusually severe the limit of speed was set by that factor combined with the limit of sending voltage which was usually placed at about 50 volts by extreme concern for the safety of the cable insulation.

It was by no means obvious which of these several factors should be considered in the electrical design of a loaded cable. With the vacuum-tube amplifier available to amplify the weak received signal to the degree necessary to operate recording mechanisms, there was no practical limit to the sensitiveness of receiving apparatus. It was, however, necessary to consider distortion as a possible limit to speed.

[3] O. E. Buckley, *Jour. A. I. E. E.*, Vol. XLIV, pp. 821–829, August 1925, *B. S. T. J.*, Vol. IV, No. 3, pp. 355–374, July 1925; J. J. Gilbert, *B. S. T. J.*, July 1927.

It is interesting to note in this connection that, though, in most previous proposals to load long telegraph cables, loading had been advocated primarily as a means of reducing distortion, practical consideration of the problem uncovered new types of distortion which were absent in the non-loaded cable. The nature of distortion of signals by a non-loaded cable was well understood, the problem having been solved long ago by Lord Kelvin. The distortion of a loaded cable is a much more complex affair since there are involved in it not only the effects of distributed inductance, capacity, resistance and leakance of the ideal cable for which the distortion is readily calculable, but also the factors of change of inductance and resistance with frequency and current, and the effects of magnetic hysteresis which are unavoidable in a practical loaded cable. Though the effect of these factors on distortion could be approximated by theoretical analysis it was considered necessary to have experimental proof that a signal could be restored in shape after passing over a loaded cable and it was primarily on this account that tests were made with an artificial loaded line. These tests showed that even the distortion of a loaded cable could be corrected by using suitable terminal networks in connection with the vacuum tube amplifier.

With the factor of distortion thus eliminated there remained duplex balance, sending voltage and received interference as possible limits to the speed of the loaded cable.

Duplex balance would, of course, set the limit of speed of operation if the cable were to be operated simultaneously in two directions as is commonly done with non-loaded cables, since it would obviously be more difficult to build an artificial line electrically equivalent to a loaded cable with its variable inductance and resistance than one equivalent to a non-loaded cable in which only resistance and capacity have to be considered. Even with non-loaded cables the difficulty of balancing is so great that the double-duplex speed is usually much less than twice the possible simplex speed and with the loaded cable, which is more difficult to balance, the relative gain in traffic capacity to be obtained by duplexing is certain to be less than with non-loaded cables. On the other hand, simplex, or one-way, operation offers very great advantages especially when used in connection with automatic operation, since it disposes of the necessity for an intricate and costly artificial line and permits dividing the full traffic capacity of the cable most efficiently to accommodate the traffic it must carry, which with most transoceanic cables is usually unequal in the two directions. For these reasons it was decided to design the first loaded cable primarily to secure efficient simplex operation. Subsequent

experience has well justified this procedure for the cables which have been made.

The problem of designing a loaded cable was thus reduced to proportioning its component parts so as to secure the desired speed of operation under the conditions imposed by the limitations of sending voltage and received interference. Considerations of safety limit the sending voltage to about 50 volts, and terminal interference as ordinarily experienced requires that the received signal shall have an amplitude of a few millivolts. The risk of increasing the sending voltage to several hundred volts would not necessarily be serious but little advantage could be gained by taking this risk since, with the materials and type of construction used, higher sending voltage would involve increased hysteresis and eddy-current losses and consequently would not result in a proportionately higher received voltage. It is, however, possible to reduce the received interference by proper termination and this is of great importance in cases where the interference is severe.

The nature of cable interference and methods of reducing it have been discussed in a paper by J. J. Gilbert [4] in which is described the method which has been used to decrease the terminal interference on the loaded cables which have been laid. This method consists in using, as the earth connection for the receiving apparatus, a "balanced" sea-earth, terminating in deep water. With ordinary cables the common practice has been to provide as the earth connection a sea-earth core, similar to the main core and sheathed with it, but extending only a few miles from shore to a point where the sea-earth conductor is connected to the sheath of the cable. While this type of earth greatly reduces the interference picked up in and near the cable terminal, it does not completely eliminate it. Almost complete elimination of the effects of disturbances originating between the termination of the sea-earth core and the shore may be obtained by providing a terminal impedance between the sea end of the sea-earth conductor and the sheath of the cable. For a non-loaded cable a combination of condensers and resistances would be required to make up such a terminal impedance, but for the loaded cable a very close approximation is secured by a simple resistance of a few hundred ohms. A few hundred feet of manganin wire, insulated like the rest of the conductor and joined to the end of the sea-earth core, serves this purpose admirably. This type of construction has been used on the New York end of the New York-Horta and on all terminals of the

[4] J. J. Gilbert, *B. S. T. J.*, Vol. V, No. 3, pp. 404–417, July 1926. See also *Electrician*, Vol. 97, p. 152, August 1926.

New York-Bay Roberts-Penzance cables, on both ends of the Horta-Emden cable, and it has also been used in the loaded cables of the Pacific Cable Board.

With the maximum sending voltage determined and with the received voltage necessary to work through interference known, the cable can be designed to give the desired speed of operation. More specifically it is necessary to provide that the attenuation for frequencies essential to the formation of the signal shall be materially less than the attenuation corresponding to the ratio of the sending voltage to the interference at the receiving end. This condition can be met by establishing the attenuation of the cable for one particular frequency related to the speed of signalling. The relation between this frequency and the speed in letters per minute depends of course on the code and method of operation used. In the case of the New York-Horta cable the fundamental frequency of a series of alternate dots and dashes of the cable code, that is, one half the center hole frequency, was used as a basis for design. For this frequency a voltage attenuation of e^{-10}, corresponding to 87 TU, can be safely assumed for recorder operation under conditions of interference such as are encountered on the New York-Horta cable. With the Baudot type of code and using the most improved apparatus, that is including a synchronous vibrating relay, a voltage attenuation of $e^{-9.5}$, corresponding to 82 TU, may be assumed for the frequency resulting from assigning 1.25 cycles to a character of the Baudot code.

The computation of the attenuation of a loaded cable requires, of course, only the substitution in the ordinary telegraph equation of the specific values of inductance, capacity, resistance, leakance and frequency which apply to the particular cable in question. The method of calculation of these electrical quantities has been discussed in previous papers and need not be repeated here.

The design of the cable is thus reduced to proportioning the elements of its construction so as to obtain the most economical cable of a given attenuation at a given frequency. The thickness of insulating material is, as has been noted above, determined practically by mechanical considerations. The electrical characteristics of the insulating material are effectively limited by the quality of gutta-percha, account being taken of its dielectric leakance which is of considerable effect on the behavior of the loaded cable though usually of almost negligible effect on non-loaded cables. With the possibilities of insulating materials thus limited the problem of electrical design reduces practically to determining the size of the conductor and the composition, size and shape of the loading material.

The desirable qualities in the loading material from the electrical point of view are high initial permeability, high resistivity and constancy of permeability in the range of magnetizing forces concerned. The exact composition of permalloy which would give the best combination of these properties would, of course, be different for different cables but for practical reasons it is desirable to choose a composition which approximates the optimum for general use. Having determined on a particular alloy, the optimum size of conductor and thickness of loading material may readily be computed on the basis of its known electrical and magnetic characteristics. With the compositions of permalloy which have been used, the optimum thickness of the layer of permalloy for a long ocean cable generally lies in the range from 0.005 inch to 0.010 inch which is fortunately convenient from the mechanical point of view. If less than the optimum thickness is assumed, the inductance will be too low and the consequent required conductor diameter will be too large. On the other hand, if more than the optimum thickness is assumed, the increase of eddy-current resistance and the effect of dielectric leakance will more than offset the gain due to the increased inductance.

In determining the optimum thickness of the permalloy it is, of course, essential to include all the resistance factors which are of consequence. In addition to eddy-current resistance and the effect of dielectric leakance there are the factors of hysteresis resistance and sea-return resistance which must, in particular, be taken into account.

The effect of hysteresis on attenuation is felt only near the sending end of the cable since over most of the length of the cable the current is so small that the hysteresis is negligible. Its effect near the terminals may be calculated by the method of successive approximations which takes account of the falling off of current and the change of hysteresis resistance with current amplitude. Ordinarily the effect of hysteresis becomes negligible beyond the first one or two hundred miles from the sending terminal. Within that range it may add as much as 10 TU to the total attenuation of the cable for the high-frequency components of the signals.

By sea-return resistance is meant the resistance which is contributed by the sea water and armor wire around the core of the cable. In low-speed non-loaded cables this factor may be safely neglected since the return current of low-frequency signals spreads out through such a great area around the cable that the resistance contributed by the sea water is negligible. With the high-frequency signals of the loaded cable, however, the return current tends to concentrate in the sea water close to the cable and much of it flows in the armor wires.

The result is a loss of energy which introduces resistance in the cable circuit, this resistance being much greater than if the armor wires were absent.[5]

The relations between sent and received voltage for some of the cables which have been laid are shown by the curves in Fig. 4, in which

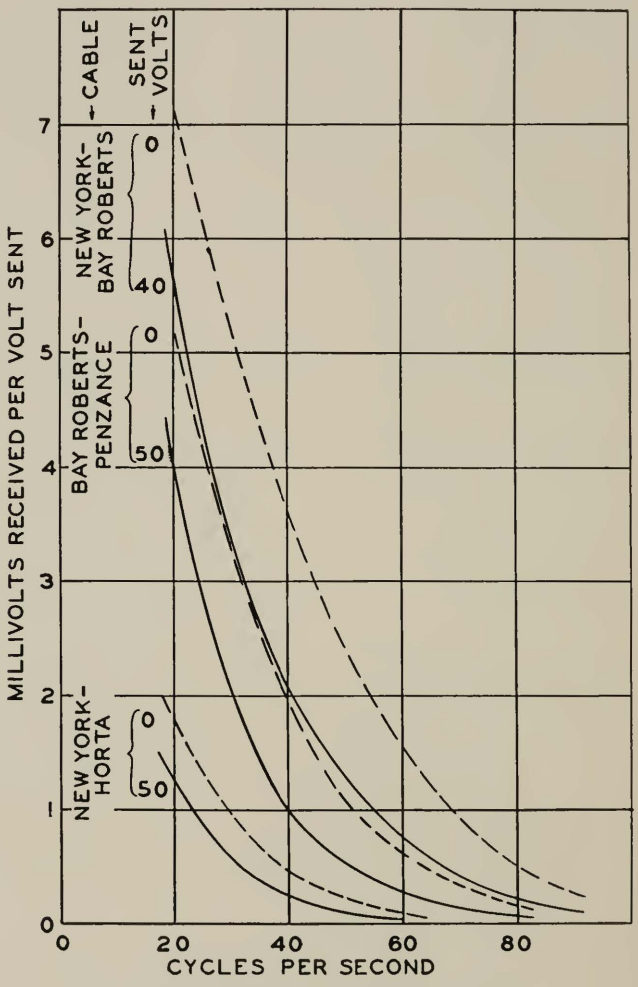


Fig. 4—Received voltage-frequency curves

the dotted curves give these ratios for zero sending voltage, obtained by extrapolation. These curves were obtained experimentally from the laid cables.

[5] See Carson and Gilbert, *Journ. Franklin Institute*, Vol. 192, pp. 705–735, 1921; *Electrician*, Vol. 88, pp. 499–500, 1922; *B. S. T. J.*, Vol. 1, pp. 88–115, July 1922.

Principles Involved in Operation

To realize practically the full benefit of the high speeds of operation of which loaded cables are capable, required the development of new types of terminal apparatus. Although many of the functions performed by the apparatus on loaded cables are similar to those involved in the operation of ordinary cables, many new problems were introduced by the higher speed of the loaded cable and by its peculiar electrical characteristics. Also new means were required to secure efficient two-way working.

With both loaded and non-loaded cables the following steps are involved in operation: translation of messages into signal impulses and the application of these impulses to the cable; correction of distortion or, as it is commonly called, signal shaping; amplification of the feeble received impulses; and reconversion of the restored received impulses into messages. The requirements to be met in accomplishing the first and last of these steps with the loaded cable are different from those in the case of the non-loaded cable, principally on account of the higher speed of the former. The requirements to be met in signal shaping and amplification are different for the loaded cable both because of its peculiar distortion and because of its high speed of operation.

The means commonly employed on non-loaded cables for sending messages involves translation of a message, usually by machine methods, into electrical impulses of the standard cable code in which a dot of the continental Morse code is represented by a positive impulse of definite duration and a dash is represented by a negative impulse of the same duration. A train of impulses of equal length but of varying polarity is thus applied to the cable at the sending end. This train of impulses is distorted and greatly attenuated by the cable but is partially restored in shape and size by terminal apparatus and is finally received on a siphon recorder which makes a record in the form of a wavy line on a paper strip. In the form and spacing of the humps and depressions of this wavy line an expert operator recognizes the positive and negative impulses which were applied at the sending end and which he is able to translate into the original message.

The necessary correction of distortion of signals on ordinary cables is accomplished by simple electrical networks at the terminals. Advantage is also taken of the mechanical characteristics of moving coil instruments. The fundamental principles [6] involved may be

[6] For a more detailed discussion of the principles of correction of distortion as applied to non-loaded cables see J. W. Milnor, *Jour. A. I. E. E.*, Vol. XLI, pp. 118–136, 1922.

roughly summed up as follows. For a line to transmit signals without distortion it would be necessary for all frequency components of the signals to be attenuated to the same degree and also for the delay or time-lag of transmission to be the same for all these frequencies. Legible signals can, however, be received if the attenuation of the combination of cable and apparatus increases with frequency, provided the increase in attenuation up to a certain value of frequency is not too great. Frequencies higher than this value are not required to form the received signal. For example, in the case of cable-code operation, a legible signal will be received if the attenuation at 1.5 times the fundamental or dot frequency is as much as 5 or 10 times the attenuation at the lowest frequencies involved in the signal, and if still higher frequency components are reduced to an inappreciable amplitude as a result of transmission through the system. The dots and dashes of the received signal will in this case be recorded as rounded but readily recognizable humps or depressions in the line traced on the siphon recorder strip.

Now the attenuation of the cable for the various frequency components is not uniform but increases rapidly with frequency. For example, on a particular transatlantic non-loaded cable a frequency of 2 cycles per second is received from the cable with one seventieth the amplitude which it had at the sending end, whereas for 4 cycles per second the received amplitude is one four hundredth of the sent amplitude and for 8 cycles per second it is only one five thousandth. The function of the distortion-correcting networks and apparatus is to attenuate the lower frequencies more than the higher ones so that the combination of cable and terminal apparatus will attenuate all frequencies up to a certain value approximately alike. The process of signal shaping may thus be regarded as one of attenuation equalization for a limited frequency band extending upward from zero. With the networks and apparatus employed on non-loaded cables the same means which serve approximately to equalize attenuation serve also to equalize time-lag. For frequencies higher than those required to form legible signals it is desirable to reduce the received current to as low a value as possible, since at such high frequencies the currents induced by sources of interference are usually stronger than those which belong to the signals. Accordingly the exclusion of these high frequencies makes the received signals more legible in being less affected by external disturbances.

The electrical networks for correcting distortion may be applied at either the sending or receiving end of the cable, or may be divided between the two ends. In ordinary cable practice it is common to

use a condenser in series with the cable at the sending end and to provide further means for signal shaping at the receiving end. The use of partial sending-end shaping has also been found desirable for the loaded cable though a modified circuit arrangement has been found more effective than the simple sending condenser.

Within recent years it has become common practice in the operation of cables to employ means for amplifying the received signals prior to relaying or recording them. This has been necessitated by the limited sensitivity of relays and recording instruments. Most of the amplifiers which have proved successful have been instruments of the moving-coil type in which a slight motion of the coil of a D'Arsonval galvanometer is caused to control a much larger source of power than that which is required to move the coil. Instruments of this type possess an advantage in that their mechanical inertia and stiffness may be used to assist in the processes of signal shaping and interference elimination. On account of mechanical limitations they are not, however, well adapted to operate at the high speed of the loaded cable.

Vacuum-tube amplifiers have been used to a limited extent on non-loaded cables and have many advantages over the moving-coil instruments, notably in their mechanical ruggedness and in the large amount of amplification which can readily be obtained with them. For use on loaded cables they have a further great advantage in that they have no frequency limitations within the range employed on cables and serve as well for high-speed cables as for low. By the use of suitable electrical networks in connection with the vacuum tubes the signals may be restored in shape, and interfering disturbances outside of the signal range of frequencies may be eliminated. A vacuum-tube amplifier which combines means for amplification, correction of distortion, and elimination of interference has been called a "signal-shaping amplifier."

With the combination of sending-end shaping network, loaded cable and signal-shaping amplifier, means are provided for conveying signals in the form of combinations of electrical impulses from one terminal to the other. Any type of telegraphic apparatus for converting messages into signals and reconverting signals into messages may be applied to complete the steps involved in one-way operation. None of the standard types of cable or land-line apparatus, however, are well adapted to meet the needs of commercial operation at the speed of the fastest loaded cables; to gain the full advantage permitted by the cable requires apparatus of special design. Special provision is also required to permit two-way operation.

There are two principal ways in which two-way working may be secured: messages may be sent simultaneously in the two directions or the cable may be used alternately in either direction. The first method is commonly called duplex and the second, simplex. Although, as was pointed out earlier in this discussion, the loaded cables which have been laid were designed primarily for simplex operation, it would be entirely possible to operate them duplex; but to do so would require the employment of an artificial line having nearly the same impedance as the cable over the range of frequencies involved in the signals. The speed of duplex operation would, of course, depend on the accuracy with which the artificial line could be made to balance the cable and this would be largely a matter of cost. Simplex operation, if the reversal of direction is made automatic, has much to recommend it over duplex. It does not require an expensive and complicated artificial line which would need frequent readjustment and it permits using the full speed of the cable to the best advantage to accommodate traffic. Means for reversing the direction may readily be associated with means for automatic printing operation and many of the objections to simplex working which are commonly thought of by the cable engineer do not apply when the reversal is thus made automatic.

Apparatus for the high-speed automatic operation of loaded cables has been described in recent papers by A. M. Curtis and A. A. Clokey. The Curtis paper [7] deals principally with the apparatus for signal shaping and amplification, while the Clokey paper [8] describes the special methods and apparatus for automatic printing telegraph operation. Some of the outstanding features of both classes of apparatus will be discussed in the following sections of the present paper.

Apparatus for Restoration of Signals

A typical circuit diagram of a loaded cable with its terminal networks for signal shaping and amplification is shown in Fig. 5. For the sake of simplicity the circuit details required for two-way operation have been omitted. Such a circuit arrangement applied to a transoceanic permalloy-loaded cable serves to connect a telegraph transmitting instrument with a receiving or recording instrument for one-way operation nearly as effectively as they could be connected by an overland telegraph line.

[7] "The Application of Vacuum Tube Amplifiers to Submarine Telegraph Cables," *B. S. T. J.*, July 1927.

[8] "Automatic Printing Equipment for Long Loaded Submarine Telegraph Cables," *B. S. T. J.*, July 1927.

At the sending end in place of the usual sending condenser there is employed the network N_1 shown in the figure. The condenser C_s may have a capacity of from 30 to 80 microfarads. It is shunted by a resistance R_1 of several thousand ohms. The resistance R_2 connecting the sending end of the cable to earth may be of the order of 100 ohms, and serves approximately to equalize the input impedance of the system over the important range of frequencies. The desirability of the resistance R_2 is peculiar to the loaded cable and is occasioned by the manner in which its characteristic impedance varies with frequency.

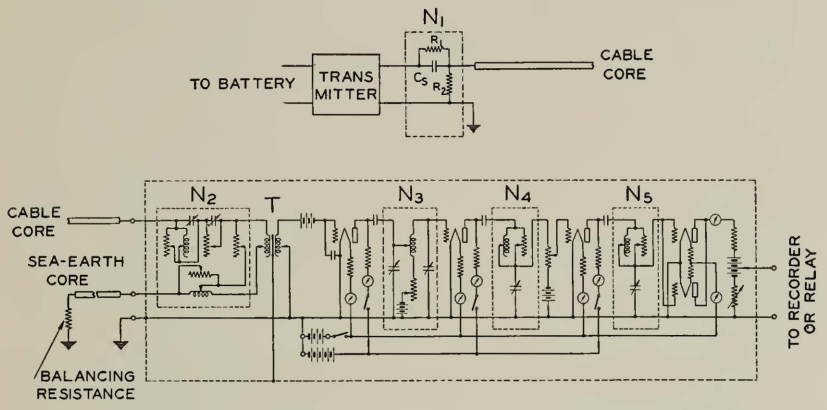


Fig. 5—Terminal networks for signal shaping and amplification

Other sending-end circuit arrangements can, of course, be used and networks combining inductances with capacities and resistances have been effectively employed. The sending-end shaping network may even be dispensed with entirely and all of the shaping done at the receiving end. There are, however, certain conditions under which this leads to the production of distortion due to hysteresis in the magnetic material of the cable and in general it is preferable to reduce the current flowing into the cable by employing sending-end shaping networks which reduce the amplitude of the low-frequency components of the signal.

The circuits employed at the receiving end for completing the process of signal shaping and for amplifying the signals may conveniently be considered in three parts, the receiving shaping network N_2, the shielded transformer T, and the amplifier which includes the interstage shaping networks N_3, N_4 and N_5.

The receiving network N_2 provides means for correction of a considerable part of the distortion introduced by the cable and in so

doing reduces the peak voltage which is applied by the signals to the primary of the transformer T and also the peak voltage which is applied to the grid of the first vacuum tube. By insertion of this shaping network between the cable and the transformer, overloading and consequent distortion are prevented.

The transformer T permits insulating the amplifier and its batteries from the cable and thereby allows the amplifier to be connected directly to earth [9] and to be effectively shielded from local electrical disturbances. Without the transformer or other means to insulate the amplifier from the cable it would be impossible to use an earthed amplifier and at the same time to secure the advantage of the balanced sea-earth in eliminating interference. The requirements for this transformer are very severe since it must be effective for frequencies as low as 0.2 cycle per second and at the same time must be constructed so that it will not pick up the external electrical disturbances generally prevalent in cable stations. The use of a permalloy core and a permalloy shield has made it possible to meet these requirements in an instrument occupying less than one third of a cubic foot.

Connected between the successive stages of the amplifier are the signal-shaping networks N_3, N_4 and N_5. These networks serve both to adjust the shape of the signal and to reduce the effects of interference outside of the signal range. Considerable advantage is gained from the fact that there are in the entire system five signal-shaping networks, each separated from its neighbors by either the cable or the vacuum tubes. This arrangement permits independent adjustment of the separate networks with very little interaction between them and greatly facilitates the systematic correction of signal shape.

The values of the various resistances, inductances and capacities in the networks at the receiving end depend, of course, on the cable as well as on the type of telegraph apparatus employed; for this reason most of the important circuit elements are made adjustable. The adjustments are made by trial, but in spite of the apparent complexity of the networks, which are more elaborate than would be required for any given cable with fixed operating requirements, the adjustments necessary to adapt the apparatus to any particular conditions can be made quite systematically. After the shaping adjustments required for a particular cable have been worked out, which usually takes not more than a few days, the amplifier can be adjusted for any speed in the range of the cable in a few minutes.

[9] The earth connection for the amplifier is preferably made to a short "sea-earth" conductor terminated on the cable sheath at a few miles from shore. The same earth conductor may be used for a transmitting earth.

The output of the amplifier may be applied to a siphon recorder or to relays, as desired, and the amount of amplification may be adjusted over a wide range to meet the requirements of any particular case. In general the power amplification needed for automatic operation of a loaded cable at its maximum speed is of the order of 10,000,000 times, which corresponds to 70 TU.

The external appearance of the signal-shaping amplifier is shown in Fig. 6. All of the receiving circuit elements shown in Fig. 5



Fig. 6—Signal-shaping amplifier

are contained in its shielded case which is made of ample size so that all of the essential apparatus units within it are readily accessible. Great care has been used in the design and construction of the amplifier unit to protect the circuit elements within it from moisture and to prevent leakage or electrostatic coupling. The output terminals of the amplifier may be connected directly to a siphon recorder or to a suitable relay. However, when the amplifier is used for the operation of relays and multiplex printing telegraph apparatus, there is associated with it an additional piece of apparatus called the relay control desk

which is shown in Fig. 7. In this unit is provided means for control and adjustment of the relays and also means to compensate the type of signal distortion commonly described as the "wandering zero" which results from the inability of the system as shown in Fig. 5 to transmit direct current.



Fig. 7—Relay control desk

Amplifiers of the type described are in commercial operation at all terminals of the Western Union and Deutsch Atlantische loaded cables. In this extensive commercial use they have been shown to require considerably less maintenance than the moving-coil instruments which are commonly used on non-loaded cables, and in fact the loss of time in operation due to troubles in the amplifiers has been almost entirely negligible. It is of interest to note that it has been possible

on numerous occasions to operate cables with these amplifiers during the entire course of severe thunder storms with the loss of only an occasional letter due to lightning discharges.

Apparatus for Automatic Operation

The first operating tests of the New York-Azores cable were made with the signal-shaping amplifier described in the preceding section. For these tests cable-code operation with a siphon recorder was employed, this type of operation being chosen because it would

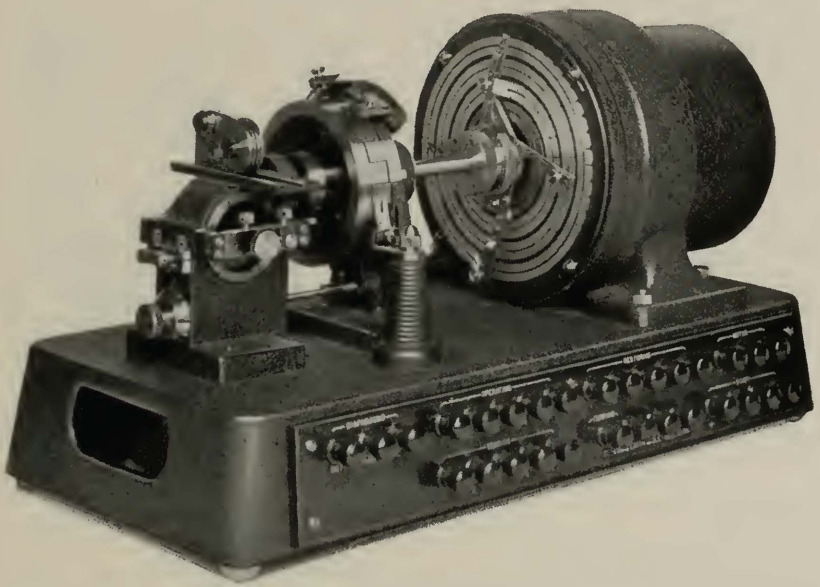


Fig. 8—High-speed cable-code transmitter

permit direct comparison of the behavior of the loaded cable with that of ordinary cables. The ordinary cable-code transmitters and siphon recorders were, however, incapable of operation at the predicted speed of over 1500 letters per minute and a new transmitter and recorder had to be provided for testing and demonstrating the operation of the new cable.

The high-speed transmitter which was developed for these tests is shown in Fig. 8. This transmitter makes use of the ordinary perforated tape used with standard types of cable transmitters but instead of opening and closing contacts by mechanical means it employs pneumatic means for this purpose, the perforated transmitting tape being utilized in the manner of the perforated sheet in a player-piano.

A commutator and relays associated with the pneumatic apparatus serve to equalize the lengths of the transmitted signals and to provide any desired ratio of "marking" to "spacing." This transmitter is capable of operating at speeds up to about 2500 letters per minute.

The high-speed siphon recorder is shown in Fig. 9. It differs from the standard instrument in many respects. A very light moving coil is supported horizontally in the strong field of an electromagnet by a bifilar suspension. A very light rigid arm attached to the coil carries a siphon pen only about 2 cm. long which writes on ordinary

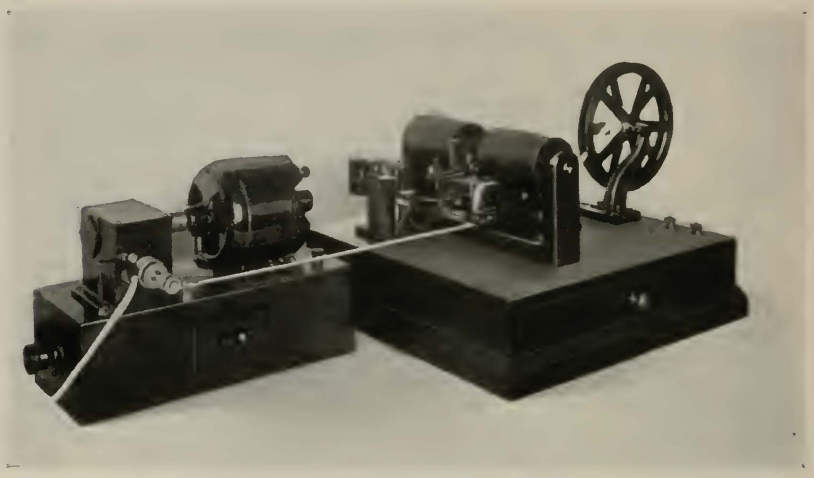


Fig. 9—High-speed siphon recorder

recorder tape drawn rapidly over a vertical table. This instrument may also be operated at 2500 letters per minute with cable-code and makes a record similar to that of the standard siphon recorder.

Both of these instruments and the signal-shaping amplifier were provided in advance of laying the first permalloy-loaded cable and were used on the first tests. A record of an early test message made on the New York-Horta cable at a speed of 1920 letters per minute is shown in Fig. 10.

Since this first cable terminated at the Azores Islands where there was no immediate demand for the full speed of which the cable was capable, the first commercial operation was conducted at a speed of only about 800 letters per minute. This was obtained with a standard cable-code transmitter and a standard type of recorder used with the signal-shaping amplifier. The cable was operated alternately in the two directions as required to accommodate traffic, the reversal of

direction of operation being controlled manually. While this type of operation served well to carry the limited traffic then available, it was not suited for efficient operation of the cable at its maximum speed, both because of the practical difficulty of dividing the rapidly received recorder tape among the three or more operators who would be required to translate it, and because of the delays resulting from manual control of reversal of direction.

To make efficient use of a high-speed telegraph cable requires some means of adapting it to the practical limitations of machines and operators, preferably by the provision of a number of separate channels of operation, each of which may be worked at a speed con-

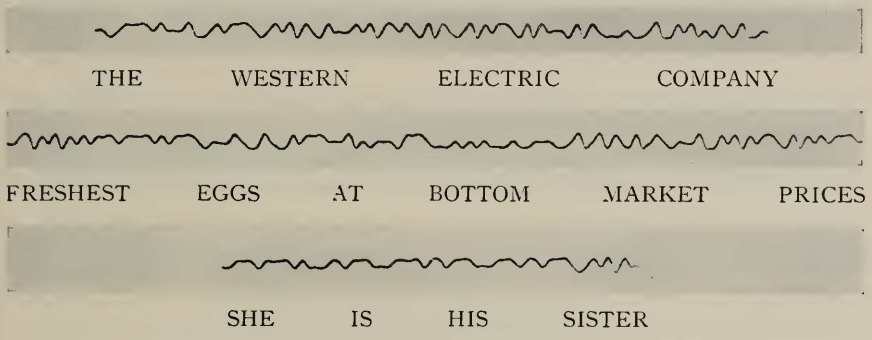


Fig. 10—Test message transmitted over New York-Horta cable at a speed of 1920 letters per minute, Nov. 14, 1924

sistent with the pace of a single operator at each end of the cable. With such multi-channel operation it is obviously necessary to provide means for either simultaneous two-way working or automatic means for direction reversal which shall not interfere with the independent operation of the several channels. Also it is very desirable to provide for automatically printing the received messages.

There are two principal methods which have been used to secure multi-channel operation with a single telegraph line—the carrier current method and the multiplex distributor method. By the former the separate channels are obtained by the modulation of separate carrier frequencies in accordance with the telegraphic signals, the line being simultaneously shared by all the channels; by the latter the line is passed in rotation from one channel to the next so that the line time is in effect divided equally among the several channels. Either method or a combination of the two can be applied to a loaded cable. The carrier current method has for several years been used on the loaded cables of the Cuban-American Telephone Co. between

Key West and Havana and has also been used on some non-loaded cables and quite extensively on land lines. The multiplex distributor method is used widely on land lines and has also been used to some extent on non-loaded cables. Of the two the multiplex distributor method makes more effective use of the line when the frequency-range is limited to about 100 cycles per second or less and the carrier current method is more effective when a considerably wider frequency-range is available. Since the frequency-range provided by the New York-Horta cable extended to about 60 cycles per second, the multiplex distributor method was the more effective means for providing multiple channels on this cable and was accordingly adopted.

With the multiplex distributor method of separating channels several different systems of operation employing different signal codes are possible and several different codes have been practically applied. Among these are the cable-code, the three-unit three-element code and the five-unit two-element or Baudot-type code. To determine which of the several possible systems can give the greater speed of operation is an extremely complex problem since it requires consideration not only of the number of characters or letters and their frequency of occurrence in messages but also of the line characteristics and the nature of interference. From the practical point of view, however, the multiplex system, which employes a code of the Baudot type, has the great advantage of availability of perfected transmitting and printing apparatus and, in view of this advantage, there seems little doubt of this being the best system for the immediate practical realization of the possibilities of a loaded transoceanic cable. In this system the line-time is divided into as many parts as there are channels of communication and each of these parts is divided into five units. The line is thus used in effect to transmit five successive signal units of either positive or negative polarity from one transmitter to its corresponding receiver, thereby sending one letter or character over one channel. It is next used to send similarly another letter on another channel and so on until a letter has been sent over each channel, whereupon a second letter is started over the first channel.

Although multiplex distributors for land lines had long been available, the standard apparatus was not suitable for realization of the full advantage of the permalloy-loaded cable. This was appreciated from the first, and long before the manufacture of a loaded cable was started the development of a system for operating it was undertaken. In several important respects the apparatus developed for the cable is different from that used on land lines.

Two-way operation is provided by automatic reversal of the direction

of sending. This is accomplished by driving from the multiplex distributor a reversing mechanism which switches the cable from sending eastward to sending westward or vice-versa at regular intervals without the loss or mutilation of a character on any channel. To adapt the apparatus to the demands for traffic, the intervals of reversal are made capable of variation over a considerable range so that the system can be used, for example, alternately one minute eastward and ten minutes westward or three minutes eastward and three minutes westward, only about five seconds being lost at each reversal.

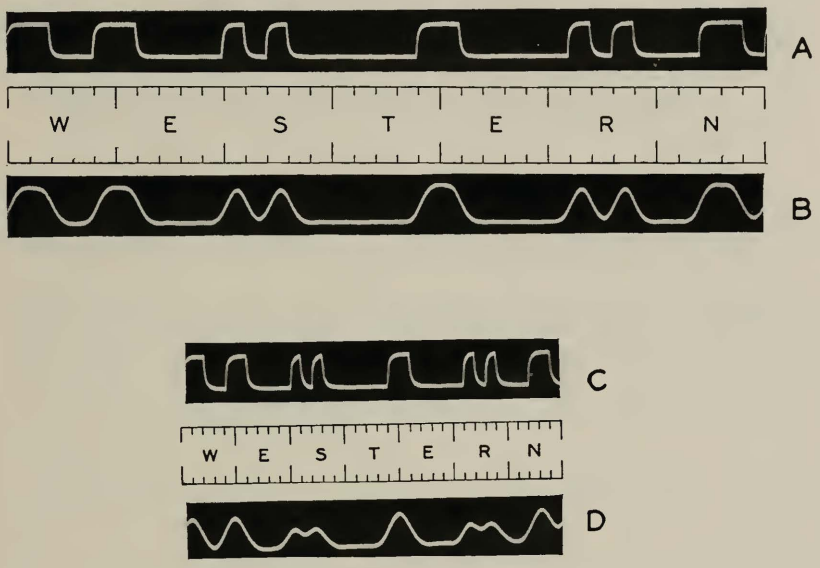


Fig. 11—Sent and received signals. *A*, Signal, sent at speed suitable for plain relay operation; *B*, Received signal shaped for simple relay; *C*, Signal sent at twice speed of *A*; *D*, Received signal shaped for vibrating relay.

(These records were made in the laboratory with an artificial line and accordingly do not show the interference which would be present in the case of a cable operated at maximum speed.)

To secure the maximum speed of operation, use is made of the "synchronous vibrating relay," a method of signal restoration developed in the course of our laboratory studies of apparatus for loaded cables. The synchronous vibrating relay takes advantage of the principle of the Gulstad vibrating relay, which has been extensively used on both land-lines and cables, but possesses a further advantage in that use is made of the synchronous multiplex distributor to secure the most effective application of this principle.

To describe and explain the circuits and apparatus of the syn-

chronous vibrating relay would be beyond the scope of this paper but the way in which it permits an advantage in speed of operation may readily be appreciated from a consideration of the signals shown in Fig. 11.

Consider first the conditions in plain relay operation without the use of the vibrating relay principle. The signal train *A*, Fig. 11, represents the word "western" as translated into the code used in the multiplex printing telegraph system. If this word is transmitted over the combined cable and distortion-correcting networks at a suitable speed for plain relay operation, it will be received in the form, *B*, in which a transmitted impulse of unit length has resulted in a received impulse of about the same amplitude as that of impulses two or more units long. A simple relay operated by the signal train, *B*, will substantially reproduce the original transmitted train, *A*.

Consider now the signal train, *C*, in which the same word "western" is transmitted at twice the speed of *A*. With the same adjustment of cable and terminal networks it will be received in the form, *D*, in which impulses of two units of length are received with the same amplitude as that with which the unit length impulse was received in *B*, whereas the amplitude of a succession of received reversals of unit length in *D* is reduced nearly to zero. Obviously, if *D* were applied to a simple relay, it would not cause the original signal train, *C*, to be reproduced. However, *C* can be reproduced from *D* by means of the synchronous vibrating relay which is arranged to supply impulses of unit length locally, unless prohibited from so doing by currents due to impulses of two or more units of length. One may regard the cable and terminal networks as converting the transmitted two-element (plus and minus) signals into three-element (plus, zero and minus) signals which the vibrating relay reconverts into two-element signals, and in this way permits operation at a speed which is much higher than is possible with a plain relay. With the Gulstad relay or with minor modifications of it, the locally interpolated impulses are supplied from a local vibrating circuit and do not always occur at exactly the right time to be most effective. With the synchronous vibrating relay, these impulses are controlled by the distributor and are therefore introduced at precisely the right time. It is interesting to note that this can be done in a system in which the incoming signals control the rate of the distributor.

Another feature of the apparatus for the loaded cable is its high degree of precision and refinement. The cost of a cable relative to that of even the most refined apparatus is so great that no considerable sacrifice of speed can be justified by ordinary economies in apparatus.

Accordingly, the efficiency to be gained by extreme precision has been sought, and to achieve this desired precision has required radical departure from the design used in land line apparatus. A photograph of one of the distributors used on the New York-Horta-Emden line is shown in Fig. 12. On this line three 5-channel distributors are used, one each at New York, Horta and Emden. Within a few seconds after one of five operators at New York prepares a perforated

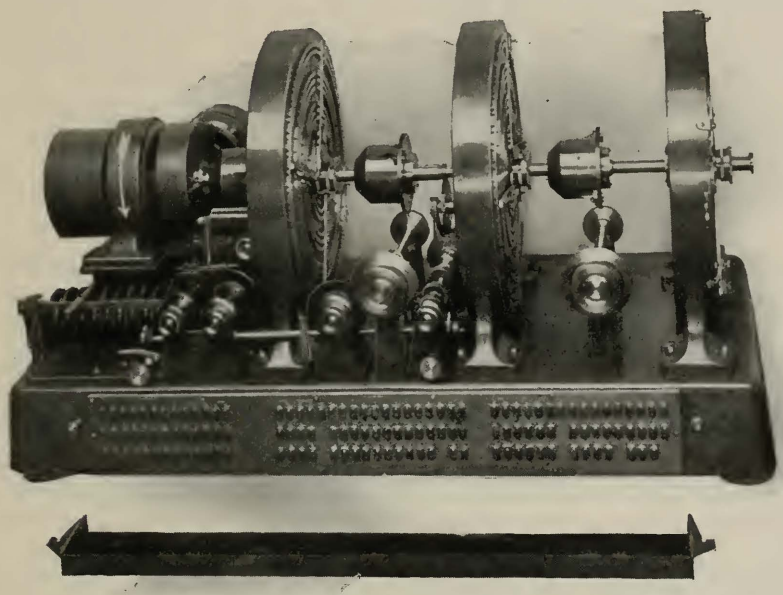


Fig. 12—Multiplex distributor used on New York-Horta-Emden line

strip on a machine resembling a typewriter, the message appears in typewritten form in Emden on a strip ready for delivery or retransmission over a land line.

While the system which I have roughly described is one which was developed principally with regard to use on a particular cable, the principal features embodied in it are applicable to any long loaded cable of the type discussed in this paper. The details of the apparatus which should be used on any cable are of course dependent on the particular requirements to be met and each installation must be engineered for its special needs if the full benefit of loading is to be realized.

Electrical Measurements of Loaded Cables

To check the assumptions made in the design of the first cables, and to obtain the information necessary for the design of the ultimate operating equipment, extensive electrical measurements were made on the three Western Union cables after they had been laid. From an analysis of these measurements the several electrical parameters of the cables were determined. To do this required new apparatus and methods, the development of which was by no means a small part of the total effort involved in the first project. Since a review of some of the methods of measurement has been given in a recent paper by J. J. Gilbert,[10] the present discussion will be limited to the apparatus and methods which seem to be of particular interest.

One of the most important tools in all our investigations concerned with the permalloy-loaded cable was the string oscillograph shown in Fig. 13. From the start it was recognized that an instrument would be needed which would give an accurate record of the manner in which the currents and voltages which were being studied changed with time and, in fact, the first step in the experimental investigation of the cable problem was to search for a suitable oscillograph. Fortunately it was not necessary to look far. A string oscillograph which had been developed for sound-ranging of artillery during the World War was quickly modified and devoted to more peaceful purposes. The present instrument differs in many details from the original but retains the invaluable asset of ability to give almost instantaneously, completely developed and fixed, a distortionless picture of a wave involving any frequencies up to about 300 cycles per second. In a study like that of signal shaping, involving the determination of the effect of numerous slight changes in adjustment of the apparatus, the advantage of such an instrument is obvious. This instrument was used both in the early studies of signal correction with the laboratory artificial cable and in the later measurements on the laid cables, and today is a useful adjunct in cable stations where it serves to show the character of the cable signals at any stage of their conversion into impulses for the recording instruments. A feature of particular value in studying phenomena such as extraneous interference on cables is that the oscillograph permits taking a continuous record over a period of several minutes.

Other special instruments which I shall only mention, but which were developed especially for the cable experiments, were an attenuation meter and a low-frequency vacuum-tube oscillator to give

[10] "Determination of Electrical Characteristics of Loaded Telegraph Cables," *B. S. T. J.*, July 1927.

voltages of constant frequency as low as 2.5 cycles per second and of very pure wave form.

To determine the various electrical parameters of the laid cable wholly from measurements at its terminals was practically impossible

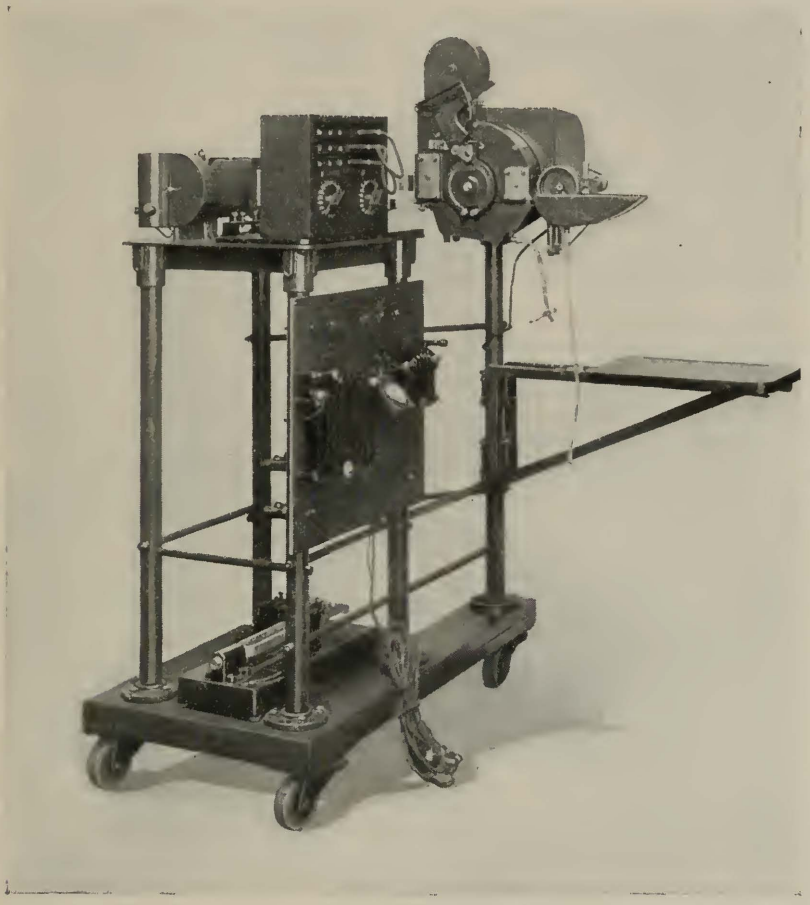


Fig. 13—String oscillograph

on account of the complicated manner in which the resistance and also to some degree the inductance and leakance vary with frequency. However, by combining the results of factory measurements with the results of measurements made at the cable terminals it was possible to determine the fundamental characteristics of the laid cable with a fair degree of accuracy. The method consists essentially in measuring, at a number of frequencies ranging from 5 cycles per

second to the highest frequency possible, the attenuation and delay or time-lag of trains of reversals transmitted over the cable.

Attenuation was measured by transmitting square-topped reversals of constant voltage and frequency and measuring the received current by means of either an amplifier and thermocouple or an amplifier and the string oscillograph, the latter method being preferable at high frequencies where the effect of interference would prohibit accurate measurement by means of the thermocouple.

The delay for steady alternating currents was measured by transmitting short trains of reversals alternately from the two ends of the cable, the transmitted and received trains at each terminal being recorded by means of the string oscillograph on a continuous strip of paper. Thus in a single cycle of operations the record at one terminal would show a transmitted train followed by a received train, while the record at the other terminal would consist of a received train followed by a transmitted train. Since a certain time is required for the establishment of the "steady state" condition at the receiving end, it was found desirable to base the measurement of the time of arrival and departure on a point well along in the train, say at the fifth to tenth cycle. The difference in elapsed time between sending and receiving at the two ends of the cable then gave twice the steady-state delay or "time of propagation" for the particular frequency for which measurements were made.

Both attenuation and delay were measured for several current values and by extrapolation to zero current the values of these quantities corresponding to a very small current amplitude could be determined. Measurements made on cores in the factory under various conditions of temperature and hydrostatic pressure gave a value of capacity for the cable and from this and the measured delay the inductance could be computed. Knowing the inductance and the capacity, the effective resistance of the cable at various frequencies could be computed from the measured values of attenuation, the value of leakance being estimated from factory measurements. This value of effective resistance should agree with the value obtained by adding together the various known components of resistance. Of the components of resistance, the copper resistance and the resistance introduced by the loading material can be computed from measurements made in the factory. The sea-return resistance can be computed from theoretical formulas and the effect of reflections from irregularities along the cable can likewise be estimated.

For the cables on which such measurements have been made, the values of effective resistance obtained from cable measurements agree

to within less than 5 per cent with the values computed as described. Part of this difference is probably due to the fact that the computed values of sea-return resistance are smaller than those actually encountered on the cable. A possible explanation of this effect is that the electrical resistivity of the earth beneath the cable is considerably higher than was assumed in the theoretical development of the formula for sea-return resistance.

A General Survey

The preceding discussion has referred principally to the progress in certain lines of development which were chosen as best suited to accomplish the result of high-speed ocean cable telegraphy. It is of interest now to consider in a general way the field of application of loaded telegraph cables and the nature of modifications which might be made in their construction and operation.

All of the loaded telegraph cables to which I have referred are relatively long. That permalloy loading should have been applied first to long cables is the natural consequence of the facts that the need for increased cable speed was principally between points far apart and that the greatest economic gain from loading could be obtained with long cables. Where high-speed cable operation was desired between points only a few hundred miles apart, it could readily be obtained with a non-loaded cable by merely making the cable large enough to give the required speed. Accordingly for short cables the operating speed was determined either by the demand for communication or by limitations of terminal apparatus. But where the necessary length was of the order of 2000 miles or more, even the heaviest cables which were considered practicable to lay and maintain were limited by the inherent characteristics of non-loaded cables to relatively low speeds, and it was accordingly for such great distances that the manifold speed advantage of the permalloy-loaded cable was of the greatest value.

To give a fair numerical estimate of the advantage of loading long cables is extremely difficult since the result depends so much on the basis of comparison and the limitations of size and operating requirements which are imposed. Probably the most nearly fair basis of comparison would be the relation of cost to speed for the old and for the new cables, but to make such a comparison requires data on cost which is forever changing. An interesting basis of comparison from the technical point of view is the ratio of the traffic capacity of a non-loaded cable operated duplex to that of a loaded cable operated

simplex, both cables being of the same length and size. The latter condition will be met if the diameter of the loaded conductor measured over the permalloy is the same as that of the copper conductor of the non-loaded cable and if the thickness of gutta-percha is the same for both. On this basis one can say that for cables of lengths from about 2000 to 3500 miles the loaded cable has approximately five times the traffic capacity of the corresponding non-loaded cable, this gain being obtained, of course, with a relatively small increase of cost.

A similar comparison might be made for shorter cables but it would have relatively little significance since in any practical case the loaded cable would probably not be made to have the greatest possible speed consistent with practicable size but would be designed with regard to the limitations of terminal apparatus or connecting lines. The problem becomes more complex as the assumed length is reduced since the shorter is the cable the greater are the number of possible ways of obtaining the desired speed and the more is the speed dependent on terminal equipment. It is, however, safe to say that, where the demand for communication is sufficiently great, loading will prove advantageous for cables of all lengths down to perhaps 100 miles or less, but for cables much less than 2000 miles long the electrical design of any particular cable will depend greatly on the use which is to be made of it.

In view of the great gain due to loading long cables it is most probable that all very long cables of the future will be loaded and it is likewise probable that long cables will be used in some cases where previously several short non-loaded sections with repeating apparatus would have been used. Loading will also be used to a considerable extent on shorter cables but it should not be expected that all of the shorter cables will be loaded since there are many cases where the demands for communication which can now be foreseen are so limited that they can be met more economically by non-loaded than by loaded cables.

In Malcolm's prediction of the loaded ocean cable, to which I have previously referred, he went so far as to suggest that even though the first loaded ocean cable would probably be of the continuously loaded type, ultimately coil-loading might be resorted to. Malcolm, of course, was not in a position to take into account the effect of such a radically new material as permalloy, and with the materials which were known to him coil-loading appeared to offer possibilities which continuous loading did not. It is interesting therefore to examine the present apparent merits of coil-loading with regard to its application to transoceanic cables.

An obvious great difficulty with coil-loaded deep-sea cables lies in the mechanical problem of laying a cable to which coils are attached or in which coils are inserted in a way to give a mechanical irregularity. Unless the coils could be made extremely small their presence would certainly interfere with passing the cable smoothly through the paying-out machinery. Cable laying and repairing are sufficiently difficult and hazardous under the most favorable conditions and any alteration in cable structure which would make these tasks more difficult is certainly to be avoided if possible. Permalloy cores for loading coils might, however, to some degree eliminate this objection to coil-loading, since with a permalloy core the loading coils may be made smaller with the result that less difficulty would be caused by the increased size of the cable at the points where the coils were inserted.

The problems of maintaining good insulation and sound joints at the loading coils are probably much more serious. Conductor joints in a cable are frequently subject to considerable stress and even with the relatively simple joints required for ordinary deep-sea cables trouble is occasionally experienced. With loading coils inserted in the cable both the coils and the joints between the coils and the core must be subject to great stress, and since the coils must be many in number to be effective, the probability of faults with even the best imaginable construction would be greatly increased.

From the electrical point of view an apparent advantage of coil-loading is that it might conceivably permit adding the required inductance without introducing so much a.c. resistance and thereby permit more closely approximating the ideal loaded cable. On the other hand, coil-loading has an electrical disadvantage which has not generally been appreciated but which is of serious practical consequence. This disadvantage lies in the distortion of signal-shape arising from the lumped character of the line. With uniform continuous loading the line is electrically smooth; such a line may introduce distortion but this distortion can be compensated for by terminal apparatus. With coil-loading the line is, in effect, a network of as many sections as there are loading coils. Such a line introduces a new type of distortion which arises in the so-called filter oscillations. Although it is theoretically possible to compensate for this effect by terminal networks, the circuits required are extremely complex and practically the limit of speed is set by the frequency of signal impulses at which filter oscillations begin to cause serious distortion. This effect can be practically eliminated by making the distances between coils sufficiently small, but as the distance between coils is diminished the otherwise possible advantages of coil-loading are likewise diminished.

Even if it could be shown that all of the apparent objections to coil-loading could be overcome, I think it is highly improbable that coil-loading would be resorted to for long deep-sea telegraph cables. Continuous loading has been given a practical trial and has proved successful and does not add greatly to the cost of a cable. Coil-loading involves risks which there is now no need to assume and its economic advantage, if any, is certainly small in proportion to the whole cost of a cable installation.

Though continuous loading as applied in several particular instances has been successful, there is no occasion to assume that the development of the art of continuous loading is completed. Modifications in continuous loading can be introduced with relatively little risk and are justified if an economic advantage can be shown. Also cable construction, apart from the loading, may be modified so as to realize more completely the advantages which loading affords. It is therefore of interest to consider some of the ways in which continuously loaded cables of the future might be different from those of the present.

Loading materials can be produced with different magnetic properties and to a limited extent the resistivity of alloys may be altered by control of composition. Higher permeability is not necessarily desirable since with increased permeability goes also increased effective resistance due to energy losses in the permalloy. In the case of any particular cable with practical limitations of dimensions, materials and costs there is an optimum permeability which, in general, is lower the higher is the frequency for which the cable is designed. Short cables designed for very high frequencies will accordingly require lower permeability than has been required for the long cables which have been made. For cables of all lengths and speeds a high degree of constancy of magnetic permeability with regard to magnetizing force is desirable. With the New York-Horta cable the inductance increases about 50 per cent when the current is increased from 0.001 to 0.1 ampere. The relative increase is less for some of the later cables, owing to improvements in loading. A loading material of high electrical resistivity is, of course, always advantageous.

There are other ways of applying continuous loading than that of Krarup. For example, the magnetic material may be electroplated onto the conductor. Such modifications may eventually come into use but the need for them does not appear to be great in the case of long deep-sea cables and there is not much economic incentive for their development. Accordingly I do not believe that changes of this type are likely to alter greatly the possibilities of submarine cables for telegraphy over long distances.

The cost and physical characteristics of insulating materials are, of course, factors of great importance. With few exceptions gutta-percha or compounds consisting mostly of gutta-percha have been used for long submarine cables. The cost of the gutta-percha insulation is a large part of the whole cost of a cable and the fact that its cost is high leads to using the least amount consistent with maintaining safe insulation. If a very much cheaper substitute or one of superior electrical properties were available, the basis of design of loaded deep-sea cables might be somewhat changed.

Even the sheath of armor wires is capable of considerable improvement as regards its effect on the behavior of a loaded cable. As pointed out previously, the sheath introduces electrical resistance due to the fact that it carries some of the return current which at high frequencies tends to concentrate around the cable. Armor wire of higher resistivity would introduce less resistance in this way or an electrical improvement might be obtained by consideration of this effect in the mechanical design of the cable sheath. At very high frequencies where the return current is closely concentrated around the cable the armor wire has an opposite effect and an improvement can be obtained by decreasing its resistance or by the addition of other conductors in parallel with it as was done on the Key West-Havana telephone cables. Some electrical resistance is introduced by the magnetic coupling between the sheath and the conductor due to the helical shape of the loading material and the armor wires. This effect is small in the cables which have been made but might be large in higher speed cables and should be taken into account in the design of such cables. A similar effect results from the use of the ordinary teredo tape on loaded cables.

With improvements in materials and construction which permit higher operating speeds and with the demand for more efficient means of handling a large volume of cable traffic, greater importance will doubtless be attached to duplex or simultaneous two-way operation, and it is of interest to consider some of the ways in which duplex operation might be secured. Of course there is no reason to assume that the loaded cables which have already been laid will not eventually be operated duplex although they were designed primarily with simplex operation in view. From the studies of both types of operation which we have made it appears more economical for the present to operate the existing transatlantic loaded cables one way at a time. Indeed this type of operation with automatic reversing apparatus possesses many advantages over the ordinary duplex methods applied to non-loaded cables. If, however, a loaded cable were designed

originally with regard to duplex operation, the possible advantage of applying duplex apparatus to it would obviously be greater than it is for any of the existing loaded cables.

There are many ways in which the design of a cable might be modified to make duplex operation more advantageous than it is on the present cables, and the problem is much too complex to permit very detailed discussion here. Improvements in constancy of inductance with variation in current and in reduction of alternating-current resistance factors would be of obvious advantage. A tapered cable with high inductance in the middle and low inductance at the ends would also have advantages in this connection and to a limited extent tapering has already been applied to the Pacific Cable Board's loaded cables by arranging the component parts of these cables during manufacture with regard to their inductance.

One of the most attractive methods of duplexing, which would also provide great flexibility of operation, is to use carrier current operation in one direction and ordinary telegraph operation in the opposite direction. Non-loaded cables are ordinarily duplexed by balancing the cable at each end with an artificial line which permits separation of the weak incoming signals from the strong outgoing ones and the limit of speed is usually set by the accuracy with which the cable may be balanced by the artificial line. By using carrier current operation in one direction and ordinary telegraphic operation in the opposite direction the incoming and outgoing signals may be separated by the combined use of artificial lines and frequency filters, in the manner long since employed on the Key West-Havana cables for carrier currents above the voice-frequency range. To design a cable for carrier current operation would, of course, require consideration of its behavior at much higher frequencies than those employed on the existing long loaded cables and would probably call for very high resistivity loading material applied in a very thin layer.

The recent spectacular development of radio both for telephony and telegraphy has raised in the minds of all the question as to whether there is any future left for ocean cable telegraphy. Opinions on this question will doubtless differ. My own opinion is that for short distances across the sea, where the demand for communication is considerable, cables always have offered more economical and satisfactory communication than radio and will probably continue to do so. For long over-sea distances I believe the cables would have faced a serious situation in competition from radio had not permalloy loading been brought forth. Now permalloy loading has so reduced that part of the total cost per word for which the cable itself is responsible

that the financial advantage of radio can never be very great. It has yet to be shown that radio telegraphy can furnish as reliable and satisfactory service as is now provided by the cables. How long the cables will continue in the leading position remains for time to tell, but it is significant that the cable companies have gone courageously ahead with new projects, and it is evident that only a much higher degree of perfection of radio communication than has yet been attained can permit wresting from the cable the advantage which it has so long maintained.

The Present Status of Wire Transmission Theory and Some of its Outstanding Problems

By JOHN R. CARSON

Synopsis: The rapid development in the technique of wire transmission and the increasing complexity of the problems involved calls for a more adequate theoretical guide and a more rigorous transmission theory. This paper gives an account, practically without mathematics, of classical transmission theory and its limitations; of the several ways the problem may be attacked more fundamentally and rigorously, and the lines along which transmission theory must be extended, as the writer has come to view the problem in the light of his own experience.

IN the present paper the term *wire transmission theory* will be understood to mean the mathematical theory of guided wave propagation along a system of parallel conductors; which is supposed to be geometrically and electrically uniform throughout its length. The theory of wave propagation along such a system is of fundamental theoretical and practical importance to the communication engineer and presents some extremely interesting and difficult problems to the mathematician. The development of the elementary or classical theory will first be briefly sketched, after which the rigorous mathematical theory will be discussed together with some of the important unsolved problems.

Historically wire transmission theory goes back to the early work of Kelvin and Heaviside. It is based on the simple idea that a transmission line (say consisting of two similar and equal wires in which equal and opposite currents flow) can be represented as consisting of uniformly distributed series inductance and resistance and shunt capacitance and leakance, these concepts deriving from electrostatics and elementary circuit theory. In accordance with this idea, if X denote the axis of propagation, the current I and voltage V are related by the familiar equations

$$\left(L\frac{d}{dt}+R\right)I = -\frac{\partial}{\partial x}V, \tag{1a}$$

$$\left(C\frac{d}{dt}+G\right)V = -\frac{\partial}{\partial x}I. \tag{1b}$$

Writing these in the usual form

$$\begin{aligned} ZI &= -\frac{\partial}{\partial x}V, \\ YV &= -\frac{\partial}{\partial x}I, \end{aligned} \tag{2}$$

where Z is the uniformly distributed series impedance and Y the shunt admittance per unit length, it is easy to show that I and V satisfy the differential equations

$$\left(\gamma^2 - \frac{\partial^2}{\partial x^2}\right) I = 0,$$
$$\left(\gamma^2 - \frac{\partial^2}{\partial x^2}\right) V = 0, \tag{3}$$

where $\gamma^2 = ZY$. The solution of these equations is

$$I = Ae^{-\gamma x} - Be^{\gamma x},$$
$$V = kAe^{-\gamma x} + kBe^{\gamma x}. \tag{4}$$

$\gamma = \sqrt{ZY}$ is called the propagation constant and $k = \sqrt{Z/Y}$ the characteristic impedance of the line. A and B are integration constants which must be so chosen as to satisfy the boundary conditions (continuity of current and potential at the line terminals). The first term represents a direct wave, the second a reflected wave, their relative values depending on terminal reflections and the terminal impressed electromotive forces.

We see therefore that in accordance with elementary or classical transmission theory, the current and potential waves are both expressible as unique simple exponentially propagated direct and reflected waves, the values of which are determined by the continuity of current and potential at the line terminals. The characteristics of the line appear only through two parameters, the propagation constant γ and the characteristic impedance k.

Generalizing the preceding, consider a system of n parallel wires, parallel to the surface of the earth. The differential equations for such a system, in terms of elementary transmission theory, are [1]

$$\sum_{k=1}^{n} Z_{jk} I_k = -\frac{\partial}{\partial x} V_j \quad (j = 1, 2 \cdots n),$$
$$\sum_{k=1}^{n} Y_{jk} V_k = -\frac{\partial}{\partial x} I_j \quad (j = 1, 2 \cdots n). \tag{5}$$

Here the physical system is represented by the parameters Z_{jk} and Y_{jk}, the Z parameters being the series impedances (self and mutual) and the Y parameters the shunt admittances. If the differential operator $\partial/\partial x$ is replaced by γ, thus confining attention to *exponentially* propagated waves, and if either the potential V or the current I is

[1] See references 9 and 10.

eliminated from (5), we get a set of n homogeneous equations in I or V, the determinant of which must vanish for a non-trivial solution. This determinant is a function of γ^2 and it has in general n roots in γ^2, indicating n possible modes of propagation, with corresponding characteristic impedances k_{jk}. The general solution is then of the form

$$\begin{aligned} I_j &= \sum A_{jk}e^{-\gamma_k x} - B_{jk}e^{\gamma_k x}, \\ V_j &= \sum k_{jk}A_{jk}e^{-\gamma_k x} + k_{jk}B_{jk}e^{\gamma_k x}. \end{aligned} \tag{6}$$

Here A_{jk}, B_{jk} are integration constants, the number of independent constants being $2n$. These are determined by the $2n$ boundary conditions at the physical terminals of the system; that is, the continuity of the n currents and n potentials. The solution represents n direct and n reflected current waves, which in general are propagated with different attenuations and different phase velocities.

The conclusions derivable from the classical theory sketched above may be summarized as follows: In a system of n parallel conductors there are in general n modes of propagation, that is, n direct and n reflected waves, which may be termed the normal modes of propagation. The distribution of the wave energy among the n modes of propagation or n component waves, is determined by the boundary conditions, which are essentially the continuity of the currents and potentials of the n wires. The system is supposed to be completely specified by the self and mutual series impedances and the self and mutual shunt admittances of the n conductors, while in the solution for the waves the physical and electrical characteristics of the line enter only through the propagation constants and corresponding characteristic impedances.

Before analyzing the theoretical basis of the preceding elementary theory, and showing its limitations, an interesting and practically important extension will be briefly touched on. The equations of the theory given above presuppose that the *impressed* electromotive forces are concentrated at the terminals of the system, and that in the line itself the electric and magnetic fields are due entirely to the currents and charges of the conductors, and consequently that the distribution of current and charge is determined entirely by their own fields. Suppose, however, that the system is in addition exposed throughout its length to an impressed field, from some disturbing source; then the preceding theory must be modified to take into account the effect of this additional field. To take the simplest case, consider a single wire parallel to the surface of the earth (ground return circuit). Let us suppose that this wire is exposed to an arbitrary field specified by an

axial electric force f at the surface of the wire and an impressed potential F (line integral of electric force to ground). The differential equations are then [2]

$$\begin{aligned} ZI &= -\frac{\partial}{\partial x} V + f, \\ YV &= -\frac{\partial}{\partial x} I - gF. \end{aligned} \tag{7}$$

(Here g is a shunt admittance. See reference 10.) Writing

$$\gamma = \sqrt{(Li\omega + R)(Ci\omega + G)},$$
$$k = \sqrt{\frac{Li\omega + R}{Ci\omega + G}},$$

this reduces to the differential equation

$$\left(\gamma^2 - \frac{\partial^2}{\partial x^2}\right) I = \frac{\gamma}{k} f + g \frac{\partial}{\partial x} F. \tag{8}$$

The solution of this equation and its practical significance have been discussed in a recent paper. The resulting analysis is of considerable practical importance in connection with the theory and design of the wave antenna and the problems of 'cross-talk' and induction.

The elementary or classical theory sketched above is essentially based on the simple concepts of electric circuit theory and its beautiful simplicity is a consequence of the fact that it is approximate only. For example the circuit parameters are only approximately calculable from the geometry of the system and its electrical constants and then only when the problem is treated as a two-dimensional one in which the variation of current and charge along the system is ignored as well as the finite velocity of propagation of their fields. Going further, it is by no means evident that even the *form* of the equations is rigorous. (We shall find that the form is rigorously valid only in an ideal case.) In the extension of elementary transmission theory, then, the first problem, as the writer sees it, is to examine the conditions under which the specification of the system by series impedance and shunt admittance parameters is justified; that is, to establish the conditions under which the classical *form* of the differential equations is valid. The second phase of this problem is to formulate a general method for calculating these circuit parameters in terms of the geometry and electrical constants of the system. The investigation of these problems leads to still further problems, arising from the fact

[2] See reference (10).

that the solutions of elementary theory, when valid, are only *particular solutions*, and therefore do not, in general, represent the complete wave.

In taking up this problem it is necessary to discard the simple concepts underlying classical transmission theory and attack the problem, *ab initio*, by aid of Maxwell's equation. Otherwise stated, our problem is to find solutions of the wave equation which satisfy the boundary conditions at the surfaces of the conductors, that is, the continuity of the tangential component of E and H, and therefore represent physically possible waves.

To put the matter otherwise, we shall place ourselves in the position of a mathematician, unacquainted with circuit theory or classical transmission theory, for whom the laws governing propagation of electromagnetic waves are formulated only by Maxwell's equations. His procedure in developing the theory of transmission along wires would be totally different from the way the theory has actually been developed. Starting with Maxwell's equations he would find that the electric and magnetic vectors satisfy a partial differential equation called the *wave equation*. He would then search for particular solutions of the wave equation which satisfy the geometry and electrical constants of the system, and therefore represent physically possible waves. The results of such a mode of approach to the problem are sketched below.

To formulate the problem concretely, consider a system of n parallel conductors, parallel to the (plane) surface of the earth, and extending along the positive X axis. The conductors may have any cross-sectional shape desired, but it is expressly assumed that they do not vary electrically or geometrically along the axis of transmission X (except at points of discontinuity or the terminals); that is to say, the transmission system is *uniform* along the axis of transmission.

Now in any medium of conductivity σ, permeability μ and dielectric constant ϵ, the electric and magnetic vectors satisfy the *wave equation*[3]

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}-\nu^2\right)F=0, \tag{9}$$

where

$$\begin{aligned}\nu^2 &= 4\pi\sigma\mu i\omega-\omega^2/v^2,\\ v &= 1/\sqrt{\epsilon\mu},\\ \omega &= 2\pi \text{ times the frequency},\\ i &= \sqrt{-1},\end{aligned} \tag{10}$$

and F may be any component electric or magnetic vector.

[3] See reference (11).

We now suppose that solutions of the type

$$F = f(y, z)e^{(i\omega t - \gamma x)} \tag{11}$$

exist, where $f(y, z)$ is a two-dimensional wave function satisfying the two-dimensional wave equation

$$\left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) f = (\nu^2 - \gamma^2) f. \tag{12}$$

In other words we search for *exponentially* propagated waves of this type; that is, waves which involve the spatial coordinate x only exponentially. It is well known that solutions of this type exist when the transmission system is uniform along the X axis.

The mathematical analysis of the problem outlined above is dealt with in detail in my paper 'The Rigorous and Approximate Theories of Electrical Transmission along Wires' (ref. 11) and the outstanding conclusions of that analysis are as follows:

The *form* of the differential equations of classical transmission theory is rigorously valid, that is, the system is specified rigorously by its self and mutual series impedances and shunt admittances, only for the ideal case of a system consisting of perfect conductors embedded in a perfect dielectric. In this case $\nu^2 - \gamma^2 = 0$ in the dielectric; $\nu^2 = \infty$ in the conductors, and the propagation constant γ is $i\omega/v$, indicating unattenuated transmission with the velocity of light, $v = 1/\sqrt{\epsilon\mu}$. The wave is a pure plane guided wave, and the electric and magnetic fields are derivable from two wave functions, one a linear function of the conductor charges and the other a linear function of the conductor currents, the determination of which, in terms of the geometry of the system, is reduced to the solution of a well-known potential problem.

Such a system, the ideal for guided wave transmission, is of course unrealizable, since there are always losses in both conductors and dielectric. For efficient transmission, however, the losses must be small and the guided wave must approximate the plane wave of the ideal case. Let us suppose, therefore, that the losses in the system are so small that *in the dielectric*, in the neighborhood of the conductors, we can set $\nu^2 - \gamma^2 = 0$, and that *in the conductors* the conductivity is so high that $\nu^2 - \gamma^2$ may be replaced by ν^2 without appreciable error. Under the circumstances where these approximations are valid it is found that the electric and magnetic fields in the dielectric and the current distribution over the cross-sections of the conductors are likewise derivable from two wave functions which are linear functions

of the conductor charges and currents respectively. The first of these is determined in terms of the geometry of the system by the solution of the same two-dimensional potential problem as in the ideal case, while the second is determined in terms of the geometry and electrical constants of the system, by a generalized two-dimensional potential problem.[4] Otherwise stated, to the approximations explained above the system may be regarded as specified by self and mutual series impedances and self and mutual shunt admittances, and these are calculable by the solution of the two-dimensional potential problems. The solution of the differential equations leads, precisely as in the classical theory, to an nth order equation in γ^2, indicating n modes of propagation. Moreover, the n corresponding waves, which will, for reasons explained below, be termed the *principal waves*, are *quasi-plane*. This means that, in the dielectric the axial electric intensity is in general small compared with the electric intensity in the plane normal to the axis of transmission; or, more broadly stated, the departure of the waves from true planarity is due entirely to dissipation in conductors and dielectric. A plane wave is here understood to mean a wave in which $E_z = H_z = 0$.

Now it is important to observe that in arriving at the foregoing result we have introduced at the outset approximations and assumptions regarding the order of magnitude of the propagation constant γ which depend on the assumption that the transmission losses are small. Fortunately these assumptions are justified, and the resulting approximate solutions are valid to a high degree of accuracy, in those systems which can be employed for the *efficient* guided transmission of electromagnetic energy; otherwise stated, the mathematical restrictions correspond to the actual requirements for efficient transmission. If, however, either the conductors or the dielectric become sufficiently imperfect, the approximations introduced and the resulting wave solutions become increasingly inaccurate and unreliable.

Suppose now that we attack the problem in a still more fundamental way: discard the assumptions regarding the order of magnitude of γ, introduced above, and attempt to deal with the problem and the solution of the wave equation in its general form. The case then is entirely different and vastly more complicated. In general, the solution can not be carried out, but a few simple systems have been studied and the results of this analysis may be generalized as follows:[5] in a system of n parallel conductors there exist, in addition to the n principal modes of propagation, an n-fold infinity of other modes of propagation,

[4] See reference (8).
[5] See reference (5).

which will be termed *complementary* modes of propagation. In general, the corresponding *complementary* waves differ from the *principal* waves in that they are not quasi-plane and are very rapidly attenuated. Consequently it appears that as regards the currents and charges, and the fields *near the conductors*, the effect of the complementary waves is usually appreciable only in the neighborhood of the physical terminals of the system so that at a distance from the terminals, usually small, they are represented with sufficient accuracy by the *principal* waves alone. At a great distance from the conductors, however, it appears that the errors resulting from ignoring the fields of the *complementary* waves may be large; in fact the complementary waves must be expressly included to take into account the phenomena of radiation.

The practical as distinguished from the theoretical importance of the foregoing resides in the fact that the principal waves corresponding to those of elementary theory represent the transmission phenomena accurately only at some distance from the physical terminals of the line and then only in the neighborhood of the wires. This defect may be of small practical consequence when the conductors all consist of wires of small cross section. When, however, conductors of large cross sections, or the ground, form part of the transmission system, the theory may be quite inadequate for some purposes. In particular, in calculating inductive disturbances in neighboring transmission systems at a considerable distance it may lead to large errors.

The discussion given above is based in part on a mathematical analysis of simple representative systems, in part on inferences from physical considerations. Unfortunately a direct frontal attack and rigorous solution of the general problem appears impossible. For example, in addition to finding the infinitely many modes of propagation the corresponding infinitely many complementary waves must be so chosen as to satisfy the boundary conditions at the physical terminals. In the classical theory these boundary conditions are simply the continuity of currents and potentials; in the rigorous formulation of the problem they are the continuity of E_y, E_z, H_y, H_z throughout the entire boundary plane ($x = 0$). Even to formulate these conditions involves specifying the impressed field throughout the plane and this is never given explicitly in technical transmission problems. While, therefore, the theory sketched above leads to inferences and conclusions of importance, the writer is convinced that some more powerful and indirect mode of attack on the problem must be devised; a rather hopeful possibility along this line will be briefly described.

As stated above, it is a reasonable inference from the general theory, that the complementary waves modify the *current* and *charge* waves appreciably only in the immediate neighborhood of the physical terminals, at least in most actual transmission systems. The essence of the method to be described consists of taking advantage of this fact and directly calculating the fields of the principal current and charge waves by means of their retarded potentials,[6] instead of employing for calculations the principal wave fields as given by the solution of the wave equation. This will now be explained in more detail.

In any transmission system energized by impressed forces introduced through terminal networks, the electromagnetic field may be analyzed as follows: (1) the impressed field, (2) the field of the terminal currents and charges, and (3) the field of the line currents and charges proper. The impressed field may be supposed to be concentrated in the terminal network, and the field of the terminal currents and charges may be supposed to be relatively unimportant except in the neighborhood of the terminals; what we are essentially concerned with is the field of the line currents and charges. Now let us suppose that we have calculated the principal wave in the system in the usual manner; corresponding to the resulting current and charge distribution, there will then be an unique corresponding field distribution determined by the solution of the wave equation, and this field is propagated in precisely the same way as the currents. But now suppose that we calculate the field of this current and charge distribution directly by means of their retarded potentials. We will find that the field so calculated is analyzable into two components: (1) a field identical with that given by the solution of the wave equation, and propagated in the same manner as the currents and charges, and (2) an additional field propagated in an entirely different way and for systems of small dissipation much more rapidly attenuated at least in the neighborhood of the conductors. We find further that the field of the principal current and charge wave does not correctly satisfy the boundary conditions at the surfaces of the conductors, which indicates that there must exist a compensating current and charge distribution. However, it appears that this compensating distribution will be relatively small and concentrated in the neighborhood of the terminals, so that we infer that its field, as calculated from its retarded potentials, can be ignored. Under such circumstances the inductive field (and the radiation field) is calculable by means of the retarded potentials in terms of the principal wave of current and charge alone.

[6] See reference (7).

To recapitulate this mode of attack, first determine the distribution of line currents and charges by means of elementary theory; that is, determine the principal wave distribution of currents and charges. Secondly, calculate the field of this current and charge distribution by means of the retarded potentials. This will give in addition to the field calculable from elementary theory an additional field the existence of which is not recognized by elementary theory. In brief, this mode of attack is based on the argument that the actual distribution of current and charge in the system is given with sufficient accuracy by elementary theory, but that in calculating the field at a distance, corrections must be introduced.

As might be expected this mode of attack presents formidable difficulties particularly when the ground plays an important rôle in the transmission phenomena. On the other hand, the analysis of a few of the simplest cases has been quite encouraging and leads one to hope that the method may at least be successfully applied to calculating the orders of magnitude of corrections which must be introduced in such important problems as, for example, inductive disturbances, in neighboring transmission systems.

The foregoing may appear to many as highly academic and theoretical. The writer's actual experience with practical transmission problems has convinced him, however, that the extension of wire transmission theory along the lines indicated above is urgently needed.

References

The papers listed below represent recent work which deals directly or indirectly with the problems discussed in the text. The relatively large number of the writer's own papers which are listed merely reflects the fact that very few specialists are working on the advanced problems of wire transmission theory.

1. "Radiation from Transmission Lines." (Carson, *Jour. A. I. E. E.*, Oct., 1921.)
2. "Radiation from Transmission Lines." (Manneback, *Trans. A. I. E. E.*, 1923.)
3. "A Generalization of the Reciprocal Theorem." (Carson, *B. S. T. J.*, July, 1924.)
4. "Das Reziprotät Theorem der drahtlosen Telegraphie." (Sommerfeld, *Jahrb. d. drahtl. Tel. u. Tel.*, 1925.)
5. "The Guided and Radiated Energy in Wire Transmission." (Carson, *Trans. A. I. E. E.*, 1924.)
6. "Über das Feld einer Unendlich langen Wechselstromdurchflossenen Einfachleitung." (Pollaczek, *E. N. T.*, 3, 1926.)
7. "Electromagnetic Theory and the Foundations of Electric Circuit Theory." (Carson, *B. S. T. J.*, Jan., 1927.)
8. "A Generalized Two-Dimensional Potential Problem." (Carson, *Bull. Am. Math. Soc.*, May–June, 1927.)
9. "Electromagnetic Waves, Guided by Parallel Wires." (Levin, *Trans. A. I. E. E.*, 1927.)
10. "Propagation of Periodic Currents over a System of Parallel Wires." (Carson and Hoyt, *B. S. T. J.*, July, 1927.)

11. "The Rigorous and Approximate Theories of Electrical Transmission along Wires." (Carson, *B. S. T. J.*, Jan., 1928.)
12. As a general reference, the treatise "Electrical and Optical Wave Motion," by Bateman, published by the Cambridge University Press, may be consulted with profit.

Appendix

The mode of attack outlined in the latter part of the text will be illustrated by an application to the simplest possible case.

Let the transmission system consist of a wire of radius a whose axis coincides with the X axis, and a coaxial cylinder of internal radius b. Both conductors are supposed to be perfectly conducting, while the dielectric in the space between ($a \leq \rho \leq b$) is supposed to be perfect. For this system we know that the principal wave is transmitted without attenuation with the velocity of light c; that is to say, $\gamma = i\omega/c$, where ω is 2π times the frequency.

We suppose that the system extends for an indefinite distance along the positive X axis so that reflected waves are absent. The principal current and charge waves are then:

$$I = I_0 e^{-i\beta x}, \qquad Q = Q_0 e^{-i\beta x}, \tag{1a}$$

where β denotes ω/c, and $i = \sqrt{-1}$. From the relation

$$\frac{1}{c}\frac{\partial}{\partial t} Q = -\frac{\partial}{\partial x} I$$

it follows that

$$Q = I. \tag{2a}$$

Now by definition the retarded potentials are

$$\Phi = \int \frac{q}{r} e^{-i\beta r} dv \quad \text{(Scalar)},$$

$$A = \int \frac{u}{r} e^{-i\beta r} dv \quad \text{(Vector)},$$

where q and u denote the charge and vector current density respectively, r is the distance between the contributing element dv and the point at which the potential is to be calculated, and the integration is extended over the entire system of currents and charges. In terms of the retarded potentials the magnetic and electric intensities E and H are given by

$$\begin{aligned} H &= \text{curl } A, \\ E &= -\text{grad } \Phi - i\beta A. \end{aligned} \tag{3a}$$

To formulate the retarded potentials of the system under consideration we have recourse to the Sommerfeld integral

$$\frac{e^{-i\beta r}}{r} = \int_0^\infty J_0(\rho\lambda)e^{-|x-x'|\sqrt{\lambda^2-\beta^2}}\frac{\lambda d\lambda}{\sqrt{\lambda^2-\beta^2}}, \tag{4a}$$

where $\rho = \sqrt{y^2+z^2}$ and J_0 is the Bessel function in the usual notation.

Applying this integral to the system of currents and charges under consideration, and remembering that they are surface currents and charges at $\rho = a$ and $\rho = b$ respectively, we get without difficulty, for $x \geqq 0$,

$$\begin{aligned}\Phi = Q_0 \int_0^\infty J_0(\rho\lambda)[J_0(a\lambda) - J_0(b\lambda)]e^{-x\sqrt{\lambda^2-\beta^2}}\frac{\lambda d\lambda}{\sqrt{\lambda^2-\beta^2}} \\ \times \int_0^x e^{x'[\sqrt{\lambda^2-\beta^2}-i\beta]}dx' \\ + Q_0 \int_0^\infty J_0(\rho\lambda)[J_0(a\lambda) - J_0(b\lambda)]e^{x\sqrt{\lambda^2-\beta^2}}\frac{\lambda d\lambda}{\sqrt{\lambda^2-\beta^2}} \\ \times \int_x^\infty e^{-x'[\sqrt{\lambda^2-\beta^2}+i\beta]}dx',\end{aligned} \tag{5a}$$

which reduces to

$$\begin{aligned}\Phi = 2Q_0 e^{-i\beta x}\int_0^\infty J_0(\rho\lambda)[J_0(a\lambda) - J_0(b\lambda)]\frac{d\lambda}{\lambda} \\ - Q_0 \int_0^\infty J_0(\rho\lambda)[J_0(a\lambda) - J_0(b\lambda)]\frac{e^{-x\sqrt{\lambda^2-\beta^2}}}{\sqrt{\lambda^2-\beta^2}-i\beta}\frac{\lambda d\lambda}{\sqrt{\lambda^2-\beta^2}}.\end{aligned} \tag{6a}$$

Since the currents are entirely axial, we have also $A_y = A_z = 0$, and from (2a)

$$A_x = \Phi. \tag{7a}$$

The first integral in Φ represents a potential wave propagated along the X axis in precisely the same way as the current and charge; it will therefore be termed the *homogeneous* potential wave. We find further that the field derivable from the *homogeneous* potentials is precisely the *principal wave* field, as given by the particular solution of Maxwell's equation, corresponding to $\gamma = i\beta$.

The second integral in Φ represents a potential wave propagated in an entirely different manner, and dying away for sufficiently large values of x. The corresponding field may be called, for want of a better term, the *heterogeneous* field, since its mode of propagation is quite different from that of the current and charge. It is this field

which represents the correction which must be added to the field of elementary theory.

It is beyond the scope of this brief appendix to discuss this solution in detail. It may be said, however, that, while the integrals representing the heterogeneous field can not be solved in finite terms, their properties can be approximately and qualitatively deduced without much difficulty.

One point of interest may be noted; the homogeneous wave is plane, that is, the axial electric intensity is everywhere zero. If we apply to the preceding formulas the relation

$$\begin{aligned} E_x &= -\frac{\partial}{\partial x}\Phi - i\beta A_x \\ &= -\left(\frac{\partial}{\partial x} + i\beta\right)\Phi, \end{aligned}$$

we get, for the heterogeneous field,

$$E_x = -Q_0 \int_0^\infty J_0(\lambda\rho)[J_0(\lambda a) - J_0(\lambda b)]e^{-x\sqrt{\lambda^2-\beta^2}} \frac{\lambda d\lambda}{\sqrt{\lambda^2 - \beta^2}}.$$

The integral term in this expression is simply the retarded potential of a ring of point sources located on the circle $x = 0$, $\rho = a$ minus the retarded potential of a corresponding ring of point charges located on the circle $x = 0$, $\rho = b$. Since this field does not vanish at the conductor surfaces $\rho = a$ and $\rho = b$, it is clear that a compensating charge and current distribution must exist.

Contemporary Advances in Physics—XV

The Classical Theory of Light, First Part

By KARL K. DARROW

FOR twenty years and more we have been hearing continually about the conflict between the corpuscular and the undulatory theories of light, and it is possible that for years to come we may be hearing about a similar contest between the wave-theory and the particle-theory of matter. Furthermore, there are intimations that if an adequate theory either of light or of matter ever is attained, it will involve conceptions of waves which in certain limiting cases approach to conceptions of particles. Already it is established that the appropriate way to attack the typical problems of the atom consists in setting up a wave-equation, and dealing with it in the same manner as one adopts to solve the typical problem of acoustics: how to determine the resonance-frequencies of a piece of elastic matter, such as a taut wire or a drumhead or a column of air in a tube. Therefore it seems opportune to restudy, with care and in detail, the great classical example of a wave-theory highly developed and widely successful—the great theory of light dimly foreshadowed by Huygens, endowed with its essential attributes by Young and Fresnel and Kirchhoff and a host of their coevals, utilized in the design of a multitude of ingenious instruments, perfected by Maxwell and connected with the theory of electricity and magnetism, and serving to this day as the basis for the theory of quanta. So doing, we shall be reminded of many triumphs of the past century of physical research, discoveries which in their time were as exciting as new quantum phenomena in ours; we shall notice certain achievements themselves as recent as those of quanta, and perhaps not less impressive; we shall retrace the reasonings which led to certain conclusions which the quantum-theories, unable to do without them and yet incompetent to derive them, have taken bodily over from their forerunner; we shall reconsider the evidence which in the litigation of a century ago caused the verdict to be rendered in favour of the wave-theory over the particle-theory; and perhaps incidentally we shall be drilling ourselves to test the evidence lately submitted and still to be submitted in the appeal of that case, and in the hearing of that other which impends.

For purposes of drill it might seem better to study the example of water-waves, which are visible; or sound-waves, of which no one denies

the existence, and no one wishes to supplant them with quanta. Ripples on the surface of a pond do furnish a precious example of wave-motion, and I presume that the notion of an undulatory theory was suggested originally by these; but it is precisely because they are visible that they fail to pose some of the questions which in dealing with light and matter are the most perplexing. Watching the leaves and the straws which float upon the surface of the water, one sees that they do not advance with the ripples; they are heaved up and down as the crests and the troughs of the wavelets pass them by. It is evident that the waves are not to be identified with the water; rather they are a form, a profile, a molding of the surface, which moves rapidly along while the substance of the liquid oscillates only a little. Now in this instance of the ripples on the pond, it is the relatively-immobile water which seems substantial and real, while that which is propagated as a wave appears to be merely a shape or a configuration, nothing more than a geometrical abstraction. It would seem strange and whimsical to assert that the liquid is a mere abstraction, but the waves are real. Yet we have to embrace this apparent absurdity in dealing with waves of light.

The example of sound-waves shows forth the paradox quite clearly. One can feel a tuning-fork; when it begins to act as a source of sound, one can see that it is quivering, and with a stroboscope one can even follow the actual course of its motion; it is even possible to see condensations and rarefactions travelling through the air, and there are numberless indirect ways of showing that sounding bodies and sound-transmitting media are matter in vibration. To the eye and to the hand, the body which vibrates is material and substantial, but not so its "vibration"—this word is only a way of saying that the shape, or the position, or the density of the body is undergoing a continuous and cyclic change.

It happens, however, that we also possess a sense for which the vibration is real but the vibrating substance is not. The ear takes no cognizance of the steel of which the tuning-fork is made, nor of the air which carries the undulations; but the ear perceives a tone. One must fully realize that the sense of hearing does not disclose that sound is vibratory. The ear does not report a sensation which goes through a cyclic variation two hundred and fifty-six times (or whatever the frequency of the fork may be) in every second. If it did, it would be perceiving the vibrating medium, not the vibration. The ear reports a sensation which is uniform, unvarying, constant; in fact, it translates a steady vibration into a constant sensation. This we have learned with ease, because of the collaboration of the other senses

which observe the bodies which oscillate. But if no one had ever felt or seen the quiverings of the humming fork, the ringing bell, or the resounding drumhead, we should be handicapped severely for discovering the true nature of sound.

Precisely so handicapped are we for discovering the nature of light. It may be that light is the vibration of a substance; but if so, the eye does not perceive that substance nor anything which fluctuates; it translates the vibration into a constant sensation. Moreover, we have no other sense which perceives that substance. When the filament of a lamp is incandescent, nothing is observed to pulsate on its surface; nothing is observed to go up and down or back and forth in the surrounding vacuum. Our instruments also fail to detect anything of which the vibrations are light. One may measure light with a photographic film or a bolometer; but the undulations—if such there be—are translated, in the one case into a steady rate of chemical change, in the other into a steady flow of electric current. In short, the eye and all our instruments register light as the ear registers tone, and not at all as the eye or the hand may register the quiverings of a sounding body; and therefore they do not report that light is vibratory. And if it be true that tangible matter is itself of the nature of a wave-motion, then the sense of touch must respond to these waves as the eye to light or the ear to sound, not reporting anything vibratory and not perceiving any medium which vibrates, but translating the vibrations into a constant sensation.

Therefore, to test whether light or matter or electricity is a wave-motion, one must make such experiments as could be made to test whether sound is a wave-motion, if there were no instrument able to perceive the vibrations of matter except the ear. Let us then suppose ourselves required to prove, to someone unable or unwilling to use any instrument except the ear, that sound is of the nature of waves; and consider how we should go about it.

The ear, we are told, is able to make distinctions of "loudness," "timbre," and "pitch." The two latter, interesting as they are, are of no immediate concern in this enterprise. It is sufficient to know that the ear makes distinctions in loudness, which according to the wave-theory correspond to distinctions in amplitude of vibration. Again, we have no concern with the exact relation between amplitude and loudness. What matters is that the former controls the latter, and therefore the latter reveals the former. Though the ear cannot detect the cyclic variations of the density and pressure of the air which make the sound, it can detect fluctuations in the amplitude of these cyclic variations. To put this statement into briefer language of

which, much later, there will be a reminiscence: the ear can detect the amplitude, but not the phase, of the vibrations. It follows, then, that we must devise tests of the wave-theory in which the amplitude of the waves shall vary from time to time, or from place to place.

Such tests are easily arranged. Let a pair of tuning-forks be set up not too close together and not too far apart. If sound consists of waves, the spherical undulations broadening outward from each of these separately must fall off steadily in amplitude as they recede, and the sound grow steadily fainter as the listener moves away. So it does; but the fact is equivocal, and cannot be taken as evidence for the waves; if the fork emitted corpuscles of sound, they would scatter apart as they flew away, and fewer would enter the ear the farther off it was placed. If, however, both of the forks are giving voice at once, and trains of spherical waves expanding outward from both, then the amplitude in the air must vary in a curious and striking manner from place to place, alternating between maxima and minima. This is just the sort of test which the ear is excellently fitted to make; being moved (or the mouth of its listening-tube being moved) from place to place in the field where the streams of sound overlap, it reports the fluctuations of loudness which are predicted from the wave-theory. By properly choosing the conditions one or more of the minima may be reduced to zero; loudness added to loudness makes silence. By properly choosing the conditions, maxima and minima may be caused to move in succession across a fixed point, listening whereat the observer hears "beats." All of these are phenomena of *interference*, and many like them are realized with light.

But it is not necessary to produce two overlapping streams of sound, in order to find evidence favouring the wave-theory. One suffices, provided that we try to separate from it a narrow jet or ray. Near one of the forks let a wall be placed, and perforated with a little hole. This seems to be an artifice for producing a constricted beam of sound proceeding like a searchlight straight outward along the line passing from the source through the hole; but it does not work that way. Instead, the tone of the fork is heard everywhere beyond the wall; sound is radiated from the hole in all directions. The aperture becomes itself a sort of secondary source, from which sound emanates sidewise as well as forward.

Precisely similar is the visible behaviour of water-waves (and incidentally of the violent compressional pulses produced in air by explosions, which have been photographed; but we are assuming that our imaginary pupil knows nothing of sound but what he hears!). Circular ripples expand over the surface of still water until one of them

meets a wall with a very narrow opening. Does a narrow segment of the ripple go clean through the opening and continue onward as a sharply-ended crescent? Not so; a new circular or semicircular ripple spreads out from the aperture as a new centre.

This is called, in the science of light, a phenomenon of *diffraction*. Actually, it is a phenomenon which reveals the law of wave-propagation—a law, which in the deceptively simple cases of spherical, circular, or infinite plane waves is artfully concealed. When one sees a circular ripple broadening over the surface of a lagoon, it seems as if each arc of the circular crest were advancing independently and of its own momentum; as if each segment of the circle at a given moment were due entirely to the corresponding segment of the smaller circle which existed a fraction of a second earlier. Nothing could be further from the truth. At a given moment, a given segment of the circle is due to the collaboration of all the segments of the earlier smaller circle; and it will collaborate with all the segments of its own circle to build the future yet larger one; and if isolated from the rest of the circumference, it would build a new family of circular ripples all by itself. Somewhat as the primitive animals which can regenerate their amputated parts, a wave-system seems to possess in each of its elements something of the power to build itself anew.

Such is the nature of ripples on water and sound-waves in air. As for a general definition of wave-motion, perhaps there is no better way of making one than to accept this manner of propagation as the distinctive mark. It may seem strange, however, that there should be any question about definition. Does not everyone know what a wave is? and is not the difference between a wave-theory and a corpuscle-theory made instantly clear by their names?

Well! it would not be hard to compile a series of paradoxical statements, by which to show that our immediate off-hand notion of a "wave" is not by any means sufficiently precise to serve as basis for an elaborate physical theory. Even in the ancient and familiar instance of circular ripples on water, even for students acquainted with the concepts of wave-length and wave-speed, there are possibilities of confusion. It is not expedient to define the wave-length as the distance from one crest to the next, for this is inconstant. It is not expedient to define the wave-speed as the speed with which a crest advances, for this may depend upon the form of the wave. It is injudicious to think exclusively about the profile of the water-surface as a sequence of visible elevations and depressions gliding steadily onward without change of shape; for any part of the profile may alter itself incessantly as it advances, departing more and more

from its original contour till it becomes unrecognizable. Definite as the wave-length and the wave-speed and the waves themselves may seem at times to be, at other times they seem indefinite and indefinable.

Now in the general theory these difficulties are removed, for the attention is focussed first of all upon an abstract entity with a neutral name—the "phase." There is a differential equation, a "wave-equation," governing the phase; and this entity is propagated in a certain very definite way conforming to the vague description which I gave above, and it has a wave-length and a wave-speed. As for the elevations and depressions of the water-surface, they copy the variations of the phase more or less faithfully, and may be computed from these; but except in particular cases the copy is not exact. To the theorist, the ripples upon the water appear as the secondary and imperfect manifestations of an abstract wave-motion, discernible only to the eye of the mind.

That the undulatory theory should introduce an abstraction even into the example whence it sprang, requiring us to imagine waves of phase underlying the tangible waves of the sea, is not at all remarkable. Often in physics a theory evolves in this way. It begins when some one notes a resemblance between two or more phenomena; it continues by the invention of a neutral and colorless mathematical expression for describing the common aspect of all these phenomena; and then the theorists take over the mathematical expression and transform and generalize and extend it, until the theory, which at first was a casual statement that two different things are in some ways much alike, eventually is defined as the entire system of solutions of some differential equation. At present there seems to be no adequate way of defining the term "wave-theory," except to say that wherever a certain differential equation is introduced and solved, there a wave-theory is adopted. Moreover, it is open to anyone to adduce new differential equations more or less like the first one, and define as a wave-theory any which involves the solutions of these. Then a wave-motion is any motion which conforms to one of the solutions; and when it comes to defining a wave, what can anyone say except that under certain restricted conditions a wave-motion may resemble a procession of ripples on water?

Such a consummation may be devoutly wished by the mathematician; to the physicist and to the expositor it is not always so welcome. As a theory increases in scope through increase of abstraction, it loses the picturesqueness which for many minds is its reason for being. One climbs and climbs, and the view indeed grows wider, but the

fascinating details of the landscape are distorted when seen from above, and finally they are lost in the haze of distance. It grows more difficult to lead others to the heights, and sometimes even the explorer cannot retrace his path and return to the firm ground of experience whence he departed. Yet repeatedly in the evolution of physics it happens that a theory, already grown so abstract that it seems almost completely severed from reality, suddenly makes new contact with the world of phenomena by a prediction so novel and daring that except for the far preliminary excursion it would probably never have been conceived; as for instance the existence of quanta, the "Einstein shift" of the lines in the spectrum of the sun, the diffraction of electrons by crystals. Remembrance of such episodes as these is an encouragement, when the path seems devious and steep.

Propagation of Waves

The laws of the propagation of waves—the so-called "laws of diffraction"—are the most important topic with which we have to deal; for they involve the very nature and definition of wave-motion, and in the end the distinction between a corpuscular and an undulatory theory of matter may rest upon these. Let us attend first of all to the making of this distinction.

Imagine, then, a multitude of particles—bullets, or atoms, or sand-grains—all rushing along through space in the same direction with the same speed, say northward with the speed c. Suppose that the location of each is stated for a certain moment of time, say t_0. The question to be asked is a very simple one, of the yes-or-no variety. At an arbitrarily-chosen point P, at an arbitrarily-chosen moment t, will there be a grain of sand or will there not?

It is easy to see what determines the answer. If at the point P at the moment t there is a grain of sand, it must have spent the time-interval extending from t_0 to t in travelling northward along the straight south-to-north path which ends at P, and therefore commences at a point P_0 due south of P and distant from it by $c(t - t_0)$ units of length. If therefore at the moment t_0 there is a grain of sand at P_0, the answer to the question is *yes*. Otherwise it is *no*. No other knowledge is required, or even relevant. It is not necessary to know the location of any of the particles which are not upon the north-south line traversing P. It is not even necessary to know the location of any particle which is upon that line, provided that at the instant t_0 it is surely somewhere else than at P_0. The state of affairs in P at t is controlled by the state of affairs in P_0 at t_0, and by nothing else whatever.

Let the same question be put in another way. Be it supposed that we are required to predict whether or not there will be a particle in the place P at the moment t; and that we are offered our choice of data concerning the places of the particles at any prior moment. Let us choose at random some point P' due south of P, distant from it by r' units of length. Then there is only one piece of information for which we have to ask: is there a sandgrain passing through P' at the moment $(t - r'/c)$, earlier than t by r'/c units of time? Any other information would be not only superfluous, but useless. Had we chosen some point lying south of P at a distance r'', the condition prevailing there at the moment $(t - r'/c)$ would have had no bearing upon the problem; but the condition there at the moment $(t - r''/c)$ would have been all-powerful. Had we chosen some point not lying south of P, nothing happening there at any time would have had any bearing upon the question to be answered.

In fine: at any moment t' there is a corresponding point P' lying south of P, which holds the destiny of the point P at the moment t. That which is predestined to befall in P at t is at every prior instant concentrated, so to speak, at a particular point of space. As time draws on towards t, this point moves on toward P, travelling always along the north-south line—travelling always, let us say, along a certain *ray* which at the proper moment carries it right into P.

All these remarks may seem too evident and trivial to be worth the making; yet they deserve attention, for it is here that the contrast lies between motion of particles and motion of waves, between undulatory theories and corpuscular. If the region around the point P is traversed not by corpuscles but by waves, it is not correct to say that the condition at the point P at the moment t is determined by the condition in some other *point* at some prior moment. Even if the waves appear to be travelling northward with the constant speed c, it is not right to say that the state of affairs prevailing in P at t is controlled entirely by the state of affairs prevailing at the moment $(t - r/c)$ at the point r units southward from P. The destiny of P at t is not travelling towards it concentrated into a point moving along a ray. Under some circumstances it appears to be right to say so; but this is only a semblance, as experiments in other conditions will clearly prove.

Suppose for instance that one is confronted with the task of sheltering the point P, first against corpuscles and then against waves, which are advancing from the south. It seems natural to put some obstacle athwart that particular north-south line which traverses P; for example, to place a solid disc so that its axis lies upon that line. If the disc can arrest all the particles which fly towards it, and cannot deflect

those which do not, then no matter how small it is it shields the point P completely against corpuscles. Not, however, against waves. Suppose that the point P is in water, where the actual waves may be seen; suppose that before the obstacle is dipped in, each of the wave-crests extends straight east and west, and they move straight northward. When the obstacle is inserted southward from P, the water at P does not become perfectly quiet. Apparently the waves curl around the edge of the obstacle, invading the zone behind it which it could have protected perfectly against corpuscles. One cannot stop a wave-motion from reaching a point merely by interrupting with some small obstruction the line along which the waves seem to be approaching it.

Now what this means is simply that, whether the obstacle be present or absent, and even though the undisturbed wave-crests move steadily due northward, the motion at P is not controlled exclusively by the motions at earlier moments at the points due south of P. To put it a little more loosely: the wave-motion at a point arrives not solely from the direction from which the wave-fronts appear to be coming, but from all directions. To put it much more strictly: imagine a sphere drawn, with any radius r, around P as centre. When we were dealing with corpuscles, we found that the state of affairs at the centre of this sphere at the moment t was entirely controlled by the state of affairs at the moment $(t - r/c)$ at one single point on the sphere (the point due south from P). Now that we are dealing with waves, we shall find that the state of affairs at the centre of the sphere at t depends upon the state of affairs all over the sphere at $(t - r/c)$. Every point upon the sphere influences the centre. Every point in the medium which the waves traverse sends forth an influence to every other point; the influence is not instantaneous, but travels from one point to another with the wave-speed c. This "influence" is often called the *wavelet.*

Too much emphasis has been laid, in the foregoing passage, upon the spheres which are centred at P; and this must now be rectified. Any closed surface whatever may be drawn around P, and the state of affairs in P at t will be determined by the state of affairs prevailing all over this surface, S, at certain prior moments t'; only, since the area-elements of S are not in general equidistant from P, the corresponding values of t' are not in general the same for all of them. The distance r, measured from P along any direction to the surface S, is in general a function of direction; consequently the time-interval, r/c, required for a wavelet to arrive at P from S along any direction is itself a function of direction; and so also is t', which is $(t - r/c)$. To every point P' on S corresponds its own value of t'; and if we know the wave-motion in every P' at its proper t', we can determine the wave-motion

in P at t. Therefore, if there is any closed surface in space, everywhere over which the wave-motion is known for all times, it is possible to compute the wave-motion at any point in the volume which that surface encloses.*

This is a feature common to all the familiar examples of wave-motion, and it is suitable for a tentative basis for a general definition of waves.

To formulate it strictly, let s be used as the symbol for any quantity which is propagated in waves. Examples of such a quantity are: the twist of a taut and twisted wire—the lateral displacement of a taut wire or a tense membrane—the excess of the pressure in the air over its average value—a component of the electric field-strength or the magnetic field-strength in a vacuum—the entirely imperceptible and hypothetical entity denoted by Ψ in wave-mechanics.

We write s as a function of x, y, z, and t:

$$s = s(x, y, z, t). \tag{1}$$

Fewer than three dimensions of space will suffice in some cases (e.g., those of the wire and the membrane); in certain problems of wave-mechanics, more than three may be required; but in dealing with sound in air and light in vacuo, three are usually necessary and sufficient. For the time being I will suppose that the speed of the waves is everywhere the same. Interesting things will happen when this assumption is discarded.

What I have loosely called "the state of affairs" in a point $P(x, y, z)$ at a moment t will involve the value of s at x, y, z, and t. Also it may involve the first and higher derivatives of s with respect to space and time, evaluated at x, y, z, t. Which of these derivatives we are required to know is something which might vary from case to case. For the present, we may consider ourselves required to know s and its first derivatives ds/dx, ds/dy, ds/dz, ds/dt.

We are to evaluate s and its derivatives at a point P at a moment t, in terms of the values which s and its derivatives possessed at certain earlier moments over a surface S enveloping P.

Let P be made the origin of our coordinate-system; let x, y, z denote the coordinates of the points on the surface S; let r denote the distance from the origin to any of these points, so that:

$$r^2 = x^2 + y^2 + z^2. \tag{2}$$

Introduce as an auxiliary the function U, defined thus:

* Naturally the surface must not be so drawn that it includes sources emitting waves during the time-interval $(t' - t)$.

$$U(x, y, z, t) = s(x, y, z, t - r/c). \tag{3}$$

The value of U in any point of the surface S at the moment t is the value of s which prevailed in that point at the moment when the "wavelet" started forth which was destined to reach the origin at t. It might be said that an observer, stationed in the origin at the moment t and inspecting the surface by means of the wavelets, observes the values of U instead of the contemporary values of s. Thus a star-gazer viewing the sky perceives, not the stars as they now are or as at some one past moment they all were, but each star separately as it was at some past epoch peculiar to itself; and the apparent arrangement of the heavenly bodies is one which in fact has never existed.

We shall be concerned not only with the value of s, but with the values of the space-derivatives ds/dx, ds/dy, ds/dz, which prevail at each point of the surface at the moment when the wavelet starts forth; for all of these will influence the value of s at the origin when the wavelet arrives there. These may be written as derivatives of U; but one must be careful here, for U is a function of x, y, z not only explicitly, but also implicitly through r; and there is a distinction to be made between total and partial derivatives, a distinction having physical importance.

To grasp this, denote by (x, y, z) the coordinates of some particular point on S, and by $(x + dx, y, z)$ those of a nearby point, and by r and $r + dr$ their respective distances from the origin, and by U and $U + dU$ the values of U in these points at the instant t. Now, U and $U + dU$ are values of s which existed *at different instants of time*, as may be seen by writing down the expressions:

$$U = s(x, y, z, t - r/c); \; U + dU = s(x + dx, y, z, t - \overline{r + dr}/c). \tag{4}$$

Therefore, if I form the total derivative dU/dx in the classical way, I am *not* obtaining the value of ds/dx which prevailed in (x, y, z) at $(t - r/c)$. To obtain this value, I must begin by subtracting the value of s prevailing in (x, y, z) at $(t - r/c)$ from the value of s prevailing in $(x + dx, y, z)$ at the same moment; that is, I must form the difference between $(U + \partial U)$ and U, meaning by the former symbol:

$$U + \partial U = s(x + dx, y, z, t - r/c). \tag{5}$$

I must then divide this difference by dx, and pass to the limit. But this is the classical way of forming the partial derivative of U with respect to x. Therefore the values of the derivatives ds/dx, ds/dy, ds/dz prevailing at the moment of departure of the wavelet which is destined to reach the origin at t, are the partial derivatives $\partial U/\partial x$,

$\partial U/\partial y$, $\partial U/\partial z$. However, the value of the derivative ds/dt prevailing at the moment when the wavelet starts is simply the derivative dU/dt, which we may as well write $\partial U/\partial t$—it makes no difference.

Our definition of wave-motion may now be stated more rigorously. A quantity s is said to be propagated by waves, if its value at the origin at the moment t is determined by the values of U, $\partial U/\partial x$, $\partial U/\partial y$, and $\partial U/\partial z$ over any surface enveloping the origin.

We now turn to another and more familiar definition of wave-motion, which shall presently be shown to fall as a special case under this one.

The Wave-Equation

There is a very celebrated differential equation of mathematical physics, known as "the wave-equation" *par excellence*. Any theory which culminates in this equation is designated as a wave-theory. The foundation of the theory of sound is the proof that the excess of pressure in the air over its average value is subject to this equation. The elastic-solid model of the luminiferous æther was partially suited to explain the phenomena of light, because the compressions and the distortions of an elastic solid conform to the wave-equation. The electromagnetic theory of light was born when Maxwell discovered interrelations between electric and magnetic fields, out of which by transformation a wave-equation could be formed. Undulatory mechanics is based upon an equation of this type which emerges during the process of setting and solving the classical equations of motion.

This wave-equation is:

$$c^2\left(\frac{d^2s}{dx^2}+\frac{d^2s}{dy^2}+\frac{d^2s}{dz^2}\right)=\frac{d^2s}{dt^2}\cdot \tag{6}$$

To demonstrate why it is called a wave-equation and what is the physical meaning of the constant c, it is customary to make a drastic simplification by assuming that the function s depends only on one co-ordinate. Such is the case, for instance, when s stands for the transverse displacement of an endlessly long taut string initially parallel to the axis of x; likewise, when it stands for the excess of the pressure of the air over its average value, and this excess is constant over every plate normal to the x-direction—a condition known as that of "plane waves." Then the wave-equation assumes the form:

$$c^2\frac{d^2s}{dx^2}=\frac{d^2s}{dt^2}\cdot \tag{7}$$

There are infinitely many solutions of this equation, and among them

are all [1] the functions of the pair of variables x and t, in which these variables appear coupled together into the linear combination $(x - ct.)$ Using f as the general symbol for a function, we may write

$$s = f(x - ct). \tag{8}$$

When such a relation prevails, any value of s which occurs at a given place x at a given moment t recurs at any other moment t' at another place x', distant from x by the length $(t' - t)/c$. *All* of the values of s existing at t are found again in the same order at t', but they have all glided along the x-direction through the same distance $(t' - t)/c$. The form, the profile, the configuration of the string are moving along with the speed c, although the substance of the string is oscillating only a little, and not even parallel to the x-direction. Now this is the property which to a certain degree of approximation ripples on water display; this in fact supplies the elementary and restricted definition of wave-motion, out of which by generalization and extension the wave-theory has grown.

Thus we see that there is reason for calling (7) a wave-equation, and identifying the constant c with the speed of the waves. Yet there are also solutions of (7) which are not of the form (8), and these do not correspond to an unchanging profile of the string travelling along with a constant speed, though by mathematical artifice they may be expressed as a summation of such; and nothing is easier than to find solutions in two or three dimensions of the general equation (6) which do not bear the least resemblance to a regular procession of converging, flat, or diverging waves. The question then arises: is there a feature common to all solutions of the "wave-equation" fitted to serve for a general definition of wave-motion?

I will now show—in the manner of Kirchhoff and Voigt—that there is such a feature, and it is precisely the one already proposed as a definition for wave-motion. If s be a function conforming to (6), and U a function related to s according to (3), then the value of s at any point at any moment is determined by the values of U and its partial derivatives $\partial U/\partial x$, $\partial U/\partial y$, $\partial U/\partial z$, over any surface surrounding that point.* The proof is long and intricate; bur for anyone who desires appreciate the nature of wave-motion, it is not superfluous.

To prove the theorem we have to manipulate the vector (call it W) of which the components are:

[1] Exceptions being made for functions which do not have derivatives, and other curiosities of the mathematicians' museum.

* The necessary requirements for continuity in s exclude sources of light from the region of integration.

$$W_x = \frac{1}{r}\frac{\partial U}{\partial x}, \qquad W_y = \frac{1}{r}\frac{\partial U}{\partial y}, \qquad W_z = \frac{1}{r}\frac{\partial U}{\partial z}. \tag{9}$$

Like U, it is a function of x, y, z not only explicitly, but also implicitly through r; we must therefore discriminate with care between total and partial derivatives. For reference, here are the formulæ [2] connecting derivatives of the one type with those of the other:

$$\frac{d}{dx} = \frac{\partial}{\partial x} + \frac{\partial r}{\partial x}\frac{\partial}{\partial r} = \frac{\partial}{\partial x} + \frac{x}{r}\frac{\partial}{\partial r} = \frac{\partial}{\partial x} + \cos(x, r)\frac{\partial}{\partial r}, \tag{10 a}$$

$$\frac{d}{dy} = \frac{\partial}{\partial y} + \frac{\partial r}{\partial y}\frac{\partial}{\partial r} = \frac{\partial}{\partial y} + \frac{y}{r}\frac{\partial}{\partial r} = \frac{\partial}{\partial y} + \cos(y, r)\frac{\partial}{\partial r}, \tag{10 b}$$

$$\frac{d}{dz} = \frac{\partial}{\partial z} + \frac{\partial r}{\partial z}\frac{\partial}{\partial r} = \frac{\partial}{\partial z} + \frac{z}{r}\frac{\partial}{\partial r} = \frac{\partial}{\partial z} + \cos(z, r)\frac{\partial}{\partial r}, \tag{10 c}$$

$$\frac{d}{dr} = \frac{\partial}{\partial r} + \frac{\partial x}{\partial r}\frac{\partial}{\partial x} + \frac{\partial y}{\partial r}\frac{\partial}{\partial y} + \frac{\partial z}{\partial r}\frac{\partial}{\partial z}$$

$$= \frac{\partial}{\partial r} + \cos(r, x)\frac{\partial}{\partial x} + \cos(r, y)\frac{\partial}{\partial y} + \cos(r, z)\frac{\partial}{\partial z}. \tag{10 d}$$

The procedure consists in forming the expression for the true divergence of W, to wit:

$$\text{div } W = \frac{dW_x}{dx} + \frac{dW_y}{dy} + \frac{dW_z}{dz}, \tag{11}$$

and integrating it over the volume comprised between two surfaces: outwardly, the surface S over which the values of U are preassigned, and which envelops the origin at which the value of s is to be computed; and inwardly, an infinitesimal sphere centred at the origin.

It will turn out that the volume-integrals of the various terms either vanish, or else may be converted into area-integrals over the two surfaces. Now the area-integral of any function f over the surface of a sphere of radius R may be written as

$$A = 4\pi R^2 \bar{f}, \tag{12}$$

in which $\bar{f}$ stands for the mean value of f over that surface—a statement

[2] In deriving the first three of these formulæ, use the relation $r^2 = x^2 + y^2 + z^2$ in evaluating $\partial r/\partial x$, $\partial r/\partial y$, $\partial r/\partial z$. In deriving the last, remember that in forming a derivative with respect to r at a point P, the increment dr is always measured along the line extending from the origin through P, for which line $x/r = \cos(x, r) = \text{const.}$; $y/r = \cos(y, r) = \text{const.}$; $z/r = \cos(z, r) = \text{const.}$; hence $\partial x/\partial r = \cos(x, r)$, etc. Or one may arrive by geometrical intuition at the formula,

$$d/dr = \cos(r, x)\,d/dx + \cos(r, y)\,d/dy + \cos(r, z)\,d/dz,$$

from which (10 d) may be obtained by means of (10 a, b, c).

which, of course, is merely the definition of f. The essential thing is that if the sphere is infinitesimal, then in the limit $\bar{f}$ becomes the value f_0 which the function f possesses at the centre of the sphere. If f_0 is finite at the origin, A vanishes in the limit; but if f varies inversely as the square of the distance from the origin, A approaches in the limit a finite value differing from zero. Upon this property our demonstration will depend.

Developing by means of (10 a, b, c) the expression given in (11) for div W, we find:

$$\begin{aligned}\operatorname{div} W = \frac{1}{r}\left(\frac{\partial^2 U}{\partial x^2}+\frac{\partial^2 U}{\partial y^2}+\frac{\partial^2 U}{\partial z^2}\right)\\ +\frac{1}{r^2}\left(x\frac{\partial^2 U}{\partial x\partial r}+y\frac{\partial^2 U}{\partial y\partial r}+z\frac{\partial^2 U}{\partial z\partial r}\right)\\ -\frac{1}{r^3}\left(x\frac{\partial U}{\partial x}+y\frac{\partial U}{\partial y}+z\frac{\partial U}{\partial z}\right).\end{aligned} \tag{13}$$

The second and third terms on the right may next be beneficially transformed by means of (10 d), using first U and then $\partial U/\partial r$ as the argument of the derivatives in that equation:

$$\frac{x}{r}\frac{\partial U}{\partial x}+\frac{y}{r}\frac{\partial U}{\partial y}+\frac{z}{r}\frac{\partial U}{\partial z}=\frac{dU}{dr}-\frac{\partial U}{\partial r}, \tag{14 a}$$

$$\frac{x}{r}\frac{\partial^2 U}{\partial x\partial r}+\frac{y}{r}\frac{\partial^2 U}{\partial y\partial r}+\frac{z}{r}\frac{\partial^2 U}{\partial z\partial r}=\frac{d}{dr}\frac{\partial U}{\partial r}-\frac{\partial^2 U}{\partial r^2}, \tag{14 b}$$

and so finally we arrive at

$$\operatorname{div} W=\frac{1}{r}\left(\frac{\partial^2 U}{\partial x^2}+\frac{\partial^2 U}{\partial y^2}+\frac{\partial^2 U}{\partial z^2}-\frac{\partial^2 U}{\partial r^2}\right)+\frac{1}{r^2}\frac{d}{dr}\left(r\frac{\partial U}{\partial r}-U\right) \tag{15}$$

as the expression to be integrated over the volume between S and the infinitesimal sphere.

Now *owing to the nature of the function U, the first term of the expression vanishes.* This is responsible for our theorem; for the volume-integrals of the remaining terms can easily be translated into surface-integrals over S and the infinitesimal sphere, from which it will follow that the value of s at the origin is determined by the values of U and its derivatives over S; but if this first term should remain, its volume-integral could not be thus transformed, and we should find that the value of s at the origin was influenced by the values of U all through the space which S encloses.

That the term in question does actually vanish is easily proved. For on the one hand it follows, from the coupling of t and r into the linear combination $(t - r/c)$ in the argument of U, that

$$\frac{\partial^2 U}{\partial t^2} = c^2 \frac{\partial^2 U}{\partial r^2}, \tag{16}$$

and on the other hand it follows, from the facts that the partial derivatives of U at any point and moment have the same values as the corresponding derivatives of s at the same point at some other moment, while the derivatives of s at *every* point and moment conform to (6)—from these it follows that

$$\frac{\partial^2 U}{\partial t^2} = c^2 \left(\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} \right). \tag{17}$$

Therefore the first term of the right-hand member of (15) is zero everywhere, and we have to perform the volume-integration only over the second:

$$\int \operatorname{div} W \, dV = \int \frac{1}{r^2} \frac{d}{dr} \left(r \frac{\partial U}{\partial r} - U \right). \tag{18}$$

Employ spherical coordinates for the integration; then the element of volume is $dV = r^2 \sin \theta d\theta d\varphi dr$, and we have:

$$\begin{aligned} \int \operatorname{div} W \, dV &= \int d\theta \sin \theta \int d\varphi \int dr \, \frac{d}{dr} \left(r \frac{\partial U}{\partial r} - U \right) \\ &= \int d\theta \sin \theta \int d\varphi \left[\left(r \frac{\partial U}{\partial r} - U \right)_S - \left(r \frac{\partial U}{\partial r} - U \right)_R \right]. \end{aligned} \tag{19}$$

This signifies that the long narrow volume-element comprised within any elementary solid angle $dw = \sin \theta d\theta d\varphi$, and limited at its two ends by the surface of the sphere and the surface S, contributes to the volume-integral the difference between the values of $(r \, \partial U/\partial r - U)$ at its two extremities. Completing the integration by considering all of these volume-elements together, we see that the volume-integral therefore becomes a pair of angle-integrals, those of the function $(r \partial U/\partial r - U)$ over the surface S and over the sphere. We may transform the first of these into an area-integral by reflecting that the elementary solid angle dw intercepts upon the surface S the area-element dS, given by the equation:

$$- dS \cos (n, r) = r^2 dw. \tag{20}$$

in which (n, r) stands for the angle between the normal to dS and the radius r drawn to dS from the origin. We must choose positive

directions for these lines. Let the radius be taken as pointing outward, and the normal as pointing inward towards the volume over which we have integrated. Then the angle is greater than 90° and not greater than 180°, its cosine is negative, and the negative sign must be prefixed to the left-hand member of (20) that dS may be positive. We make this transformation in (19), and the first of the angle-integrals becomes:

$$\int d\theta \sin\theta \int d\varphi \left(r\frac{\partial U}{\partial r} - U \right)_S = \int_S dS \cos(n, r) \left(\frac{1}{r}\frac{\partial U}{\partial r} - \frac{U}{r^2} \right). \quad (21)$$

The second of the angle-integrals in (19) relates to the infinitesimal sphere. We transform it as we did the first. Now, however, the process is more simple, for the radius r is constant and equal to R, and the angle (n, r) is zero; hence

$$dS = R^2 \sin\theta d\theta d\varphi \quad (22)$$

and the second angle-integral becomes:

$$\int d\theta \sin\theta \int d\varphi \left(r\frac{\partial U}{\partial r} - U \right)_R = \int dS \left(\frac{1}{R}\frac{\partial U}{\partial r} - \frac{U}{R^2} \right). \quad (23)$$

Here we meet the situation for which equation (12) was introduced—the integration of a function over an infinitesimal sphere. Denote by f the integrand in (23), viz.,

$$f = \frac{1}{R}\frac{\partial U}{\partial R} - \frac{U}{R^2}, \quad (24)$$

by $\bar{f}$ the mean value of f over the sphere of radius R, by U_0 the value of U at the centre of the sphere, which is the origin. As R approaches zero, the surface-integral in (23) approaches a limit A_0 which coincides with the limit approached by $4\pi R^2 f$:

$$A_0 = \operatorname*{Lim}_{R=0} 4\pi R(\overline{\partial U/\partial R}) - 4\pi U_0. \quad (25)$$

Unless the mean value of $\partial U/\partial R$ should vary as the first or a higher power of $(1/R)$—a possibility which must be guarded against—the first term on the right of (25) will vanish. Under this restriction, then, A_0 is equal to $-4\pi U_0$. Now at the origin U is identical with s, by definition (equation 3). Consequently U_0 is identical with the value of s at the origin at the moment t—the very thing which we set out to calculate. For this—let it be called s_0—we have attained the following equation:

$$4\pi s_0 = \int \operatorname{div} W dV + \int dS \cos(n, r)(\partial U/\partial r). \quad (26)$$

We still have a volume-integral in the formula; but there is a very noted theorem whereby it may be transformed with sign reversed into a surface-integral over the two surfaces, S and the sphere, which bound the region of integration. According to Gauss' Theorem, any vector function satisfying certain simple conditions of continuity throughout a region enclosed by a surface enjoys this property: its volume-integral through the region is equal to the area-integral, over the enclosing surface, of the projection of the vector upon the direction of the *outward*-pointing normal. This latter is the same in magnitude and opposite in sign to the projection upon the direction of the *inward*-pointing normal, which it is traditional to prefer. The theorem is not valid, if the vector should exhibit certain singularities within the volume; one of the reasons for introducing the infinitesimal sphere is that the vector W has a singularity at the origin, which point must therefore be excluded from the volume of integration.

Remembering the definition of W (equation 9), we see that its projection upon the direction of the *inward*-pointing normal at any place upon either surface is

$$\begin{aligned} W_n &= W\cos(n, r) = W_x\cos(n, x) + W_y\cos(n, y) + W_z\cos(n, z) \\ &= \frac{1}{r}\left[(\partial U/\partial x)\cos(n, x) + (\partial U/\partial y)\cos(n, y) + (\partial U/\partial z)\cos(n, z)\right] \\ &= \frac{1}{r}(\partial U/\partial n), \end{aligned}$$

in which $(\partial U/\partial n)$ stands for the rate at which the function U, owing to its *explicit* dependence upon x, y, and z, varies as one moves *inward* along the normal to the surface. The distinction drawn in the foregoing pages between partial and total derivatives must be remembered. The partial derivative $\partial U/\partial n$ existing at any point P and moment t is equal to the value which the corresponding derivative ds/dn of the function s possessed at that same point at the earlier moment $(t - r/c)$.

The quantity W_n is to be integrated over the surface of the sphere and over the surface S. However, the integral over the sphere vanishes as the radius of this latter approaches zero, for the same reason—and under the same restriction—as caused the integral of the first term in (25) to vanish. This leaves us with nothing but the integral of W_n over the surface S, so that eventually:

$$\int \operatorname{div} W\, dV = -\int_S \frac{1}{r}(\partial U/\partial n)dS. \tag{28}$$

$$4\pi s_0 = \int_S dS\left[\cos(n, r)\frac{\partial}{\partial r}\left(\frac{U}{r}\right) - \frac{1}{r}\frac{\partial U}{\partial n}\right]. \tag{29}$$

We have spoken of the value of s at the origin of coordinates, for mathematical convenience; but in reality the "origin" is any point P, and S is any surface enclosing that point, and s_0 is the value of s in the point P at any moment t, and U is the value of s in any point distant by r from P, evaluated at the moment $(t - r/c)$. Hence (29) may be written thus:

$$4\pi s = \int_S dS \left[\cos(n, r) \frac{\partial}{\partial r} \left(\frac{s(t - r/c)}{r} \right) - \frac{1}{r} \frac{\partial s(t - r/c)}{\partial n} \right]. \quad (30)$$

The task is achieved. It has been proved that when a function conforms to "the wave-equation" it conforms also to the first-suggested definition of a wave-motion, in that its value at any time and place is determined by its anterior values and those of its derivatives over a surface completely enclosing the place. Moreover the actual formula has been derived whereby the value at any point and moment can be computed when the values all over any surrounding surface are known at the appropriate prior moments.

Introduction of the Ideas of Frequency and Wave-length

Hitherto I have spoken chiefly of an extremely abstract "something," denoted by a symbol s, and possessed of the property that its value at any point and moment is built up out of contributions despatched at earlier moments from all of the area-elements of any continuous surface which encloses the point; these contributions being borne as it were by messengers, who travel to the point at a finite speed from the various area-elements whence they depart. Only one physical constant has been introduced, and this is the speed of these messengers. This is the constant which appears in the wave-equation (6), being there denoted by c. It is commonly called the speed of the waves; but, for various reasons which will eventually appear, it had better be called the *phase-speed*. Now there are two other constants familiar in our experience with water-waves and sound; they are *frequency* and *wave-length*. Let us try to import them into the general theory.

At any point of a water-surface over which uniform ripples are passing, the elevation is a periodic function of time; so also are the pressure and the density at any point of a gas through which uniform sound is flowing, or the displacement of either prong of a steadily-humming tuning-fork. Any periodic function of time is either a sine-function, or a composite of sine-functions. It is suitable therefore to begin by analyzing the case in which the function is a sine. Using f to denote any of the quantities above mentioned or anything behaving like them—say displacement, for example—let us write:

$$f = F \sin(nt - \delta) = F \sin \varphi. \quad (41)$$

In this very familiar form, F stands for *amplitude* and n for 2π times the *frequency*, and δ for something which is commonly called the phase; but it will be better to reserve this name for the entire argument of the sine-function:

$$\text{Phase} = \varphi = nt - \delta = \arcsin (f/F), \tag{42}$$

and I shall use it henceforth in this sense.

Now, in general, both the amplitude F and the phase φ vary from point to point—the latter because δ is often a function of position. Consequently f is a function of x, y, z and t; and it is immediately important to find out whether f satisfies the wave-equation. One who is familiar chiefly with the standard one-dimensional case of waves on a string is likely to think that this is true as a matter of course. However, on differentiating f twice with respect to x or y or z and taking due account of the dependence of both F and δ upon these variables, one sees directly that in general it is not true—not unless the functions F and δ conform to definite and sharply restrictive conditions.[3]

This appears a rather disconcerting result. However, if instead of f we envisage f/F—the value of the displacement at each point referred to its amplitude there as a unit, or the sine of the phase—it turns out that the condition under which f/F obeys the wave-equation is far less drastic. Forming the derivatives, we obtain:

$$\frac{\partial^2}{\partial t^2}\frac{f}{F} = -n^2 \sin \varphi, \tag{43}$$

$$\nabla^2 \frac{f}{F} = (\nabla^2 \varphi) \cos \varphi - \left[\left(\frac{\partial \varphi}{\partial x}\right)^2 + \left(\frac{\partial \varphi}{\partial y}\right)^2 + \left(\frac{\partial \varphi}{\partial z}\right)^2 \right] \sin \varphi. \tag{44}$$

Evidently, in order that the function f/F shall conform to the wave-equation, it suffices that

$$\nabla^2 \varphi = 0, \tag{45}$$

This condition is fulfilled by a variety of functions, including all which are linear in x, y, and z—the case of "plane waves," which as we shall see is one of those permissible when the wave-speed is everywhere the same, as I have been assuming.

Next we will evaluate the speed of the phase-waves, and incidentally we shall be led to a new aspect of wave-motion. When the phase

[3] The general expression for $\nabla^2 f$ is $(\nabla^2 F - F|\nabla\phi|^2) \sin \phi + (2\nabla F \cdot \nabla\phi + F\nabla^2\phi) \cos \phi$. The coefficient of $\cos \phi$ must vanish, if f is to satisfy the wave-equation with real phase-speed. By introducing the notion of "imaginary phase-speed" one may continue to regard the function f as conforming to the wave-equation, even though the coefficient of the cos-term does not vanish.

conforms to (45), it follows from (43) and (44) that

$$\nabla^2 \frac{f}{F} = \frac{1}{n^2}\left[\left(\frac{\partial\varphi}{\partial x}\right)^2 + \left(\frac{\partial\varphi}{\partial y}\right)^2 + \left(\frac{\partial\varphi}{\partial z}\right)^2\right]\frac{\partial^2}{\partial t^2}\frac{f}{F}. \tag{46}$$

The quantity in brackets, being the square of the magnitude of the gradient of φ, shall be denoted by the usual symbol $|\nabla\varphi|^2$. Then for the square of the phase-speed we obtain $n^2/|\nabla\varphi|^2$, and for the phase-speed itself:

$$c = = + n/|\nabla\varphi|, \tag{47}$$

The quantity n is 2π times the frequency. Also, it is the time-derivative of φ, as follows directly from the definition in (41); consequently:

$$c = = (\partial\varphi/\partial t)/|\nabla\varphi|. \tag{48}$$

This is an equation with a very important meaning, which will now be displayed.

Select any point in the medium and any moment of time t_0, and denote by φ_0 the value of φ prevailing then and there. Singular cases excepted, there is an entire surface containing the point in question everywhere over which φ has the same value φ_0. This surface is by definition a *wave-front*. Call it S_0. At a slightly later moment $t_0 + dt$, there will also be a surface everywhere over which the value of φ is φ_0. It will however not be the same surface S_0, but another—a wave-front S_1 so placed that from any point P_0 on S_0 the nearest point on S_1 is reached by measuring the length cdt along the line normal to S_0, in the sense in which φ is decreasing (the sense opposed to the gradient of φ).

To see this, imagine a particle which at the instant t_0 is travelling through P_0 along the direction normal to S_0 in the sense just stated, with a speed to be designated by u. At the instant $t_0 + dt$ it occupies a point where the value of φ then prevailing is given by the formula:

$$\begin{aligned}\varphi_0 + d\varphi &= \varphi_0 - |\nabla\varphi|ds + \left(\frac{\partial\varphi}{\partial t}\right)dt \\ &= \varphi_0 - u|\nabla\varphi|dt + \left(\frac{\partial\varphi}{\partial t}\right)dt,\end{aligned} \tag{49}$$

for in the time-interval dt it travels over a distance $ds = = udt$ along the normal to the surface S_0, and along this normal the slope of the function φ is equal to $|\nabla\varphi|$, and meanwhile at each point of space φ is varying directly with time at the rate $(\partial\varphi/\partial t)$. Now if the imagi-

nary particle happens to be moving with just the speed defined by the equation

$$u = (\partial\varphi/\partial t)/\nabla\varphi = c, \tag{50}$$

the coefficient of dt in equation (49) vanishes; that is, the particle as it moves along keeps up with the preassigned value of φ; but this is the same thing as saying that c is the speed of the wave-front.

The phase-function φ therefore possesses a quality which in itself suggests one aspect of a wave-motion. It is not periodic, neither does it conform to the wave-equation; but each of the surfaces over which φ has any constant value is perpetually travelling. Each of them may be changing continually in size, it may even be changing in shape; but each retains its identity, and if it is completely known for any given instant, its past and future history are determined completely; for each of the area-elements of such a surface is moving at the speed c and in the direction normal to itself, and from the position of each area-element at the moment t we can determine the position of the area-element into which it evolves at the moment $t + dt$, and repeat this process of prediction or retrospect *ad infinitum*. I will speak of this state of affairs as *propagation by wave-fronts*.

Having determined by (48) the relation between frequency and phase-speed, we now can give both the definition and the formula for the wave-length. The wave-length λ is by definition the quotient of phase-speed by frequency:

$$(n/2\pi)\lambda = c, \tag{51}$$

and in this special case, the formula for it is:

$$\lambda = 2\pi/|\nabla\varphi|. \tag{52}$$

The reason for giving this quantity a name, and such a name as "wave-length," arises from the best-known and too-exclusively-known special case, that of "plane waves" commonly so called—meaning not only that the wave-fronts are plane, but also that the amplitude is constant over each. Such waves travelling along any direction, say that of x, are described by the expression:

$$f = F \sin(nt - mx), \qquad F = \text{constant}. \tag{53}$$

The wave-fronts—that is to say, the surfaces over which $\varphi = (nt - mx)$ is constant at any moment—are planes normal to the axis of x. These planes are likewise the surfaces over which the displacement f is constant at any moment, and it is tempting to define the wave-fronts as the loci of constant displacement; but this is a coincidence which should be regarded as an accident. At a given moment, any value of $\sin\varphi$

which is found anywhere repeats itself at intervals $2\pi/m$ all along the x-direction; exactly as, at a given point, any value of sin φ which is found at any moment repeats itself at intervals $2\pi/n$ all through time. Owing to the coincidence aforesaid, any value of f which is found anywhere also repeats itself at spacings $2\pi/m$ along the direction of x. This constant spacing serves as the elementary definition of wave-length; and in this special case the elementary agrees with the general definition, for

$$m = |\partial\varphi/\partial x| = |\nabla\varphi| = 2\pi/\lambda. \tag{54}$$

But there is an almost equally simple case in which the spacing between wave-fronts and the spacing between surfaces of constant f are not the same. I refer to the case of spherical waves of sound or the circular ripples on a water-surface, in which we have:

$$f = F \sin (nt - mr), \qquad F = \text{constant}/r. \tag{55}$$

In this case f/F does not conform to the wave-equation, but the function f does. Nevertheless it is the phase, and not the displacement, which advances steadily outward (or inward) in a sequence of steadily diverging (or converging) spherical wave-fronts which expand or contract with the constant phase-speed c. For any two of these spherical wave-fronts differing in radius by $2\pi/m$, the values of ϕ are the same. The surfaces of constant f are also spheres, but they expand or contract with variable speed, and for any two which differ in radius by $2\pi/m$ the values of f are different. This shows that one must not be misled by experience of plane waves into defining "wave-length" as "distance between points where at the same moment the displacement is the same" but must hold fast to the phase as the central fact of any wave-motion.

If the phase-function ϕ does not vary in space, we have the case of *stationary waves*. The coefficient of cos ϕ in the general expression for $\nabla^2 f$ (footnote on p. 339) now vanishes automatically; the coefficient of sin ϕ reduces to the term $\nabla^2 F$, and this must be equated to $-n^2F/c^2$, which if n and c are preassigned leads to an alternative form of the wave-equation

$$\nabla^2 F + \frac{n^2}{c^2} F = 0 \tag{56}$$

very common in acoustics and in wave-mechanics.

In summary:

We have considered two definitions of wave-motion: *first*, that the state of affairs at any point and moment in the medium is controlled by the state of affairs at earlier moments all over any continuous sur-

face drawn in the medium completely around the point; *second*, that the function which is propagated in waves conforms to the so-called "wave-equation."

We have found that these definitions are compatible with one another, the latter being included under the former.

We have applied them to the case of a function which at any particular point of the medium varies as a sine-function of time, thus:

$$f = F(x, y, z) \cdot \sin \varphi; \qquad \varphi = nt - \delta(x, y, z),$$

and have found:

(*a*) that provided the functions F and δ conform to certain stipulations, the function f will satisfy the wave-equation;

(*b*) that φ itself is propagated by wave-fronts; although there is nothing periodic or vibratory about φ, each surface over which φ possesses any constant value wanders onward through space, changing, it may be, in shape as well as position;

(*c*) that the speed with which the wave-fronts of the phase-function φ travel is the speed at which the contributions, out of which the value of f/F at any point and moment is built up, travel to that point from any environing surface.

A Test of Kirchhoff's Theorem

Having formed the conceptions of plane waves and sine-vibrations and frequency and wave-length, we now can practice on Kirchhoff's theorem by applying it to a problem of which the answer is predetermined, so preparing ourselves for other problems of which the answers can be discovered only by means of the theorem.

Imagine plane monochromatic waves, of frequency $\nu = 2\pi n$, wave-length $\lambda = 2\pi/m$, and phase-velocity $c = \nu\lambda = n/m$, travelling in the positive x-direction in an endless procession through an infinite medium. They are described by the function:

$$s = \cos (nt - mx), \tag{61}$$

which is a solution of the wave-equation (6).

The value s_0 of the function s at the origin at the moment t is by hypothesis

$$s_0 = \cos nt \tag{62}$$

and according to Kirchoff's theorem it is given by the following equation:

$$4\pi s_0 = \int dS \left[\cos (n, r) \frac{\partial}{\partial r} \frac{U}{r} - \frac{1}{r} \frac{\partial U}{\partial n} \right], \tag{63}$$

in which the integrand on the right is integrated over any surface completely enclosing the origin. For any such surface, then, the integral on the right of (63) must be equal to $4\pi \cos nt$.

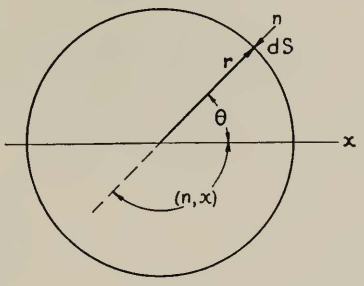


FIG. 1

This we proceed to test for the simplest of such surfaces, a sphere centred at the origin.

Denote by R the radius of the sphere; by θ, the angle between the positive direction of the x-axis and the line drawn from the origin outward through any point P of the sphere. The inward-pointing normal at P is directed oppositely to this line, and hence $\cos(n, r) = -1$ and $\cos(n, x) = -\cos\theta$. The function U and its derivatives are these:

$$\begin{aligned} U &= \cos\left[n\left(t - \frac{r}{c}\right) - mx\right] = \cos(nt - mr - mx), \\ \partial U/\partial r &= m \sin(nt - mr - mx), \\ \partial U/\partial n &= (\partial U/\partial x)\cos(n, x) = (\partial U/\partial x)\cos(\pi - \theta) \\ &= -m\cos\theta \sin(nt - mr - mx). \end{aligned} \tag{64}$$

In each of these we are to set $r = R$ and $x = R\cos\theta$ in preparation for integrating over the sphere. The element of area is

$$dS = 2\pi R^2 \sin\theta d\theta \tag{65}$$

and the limits of integration are 0 and π. Therefore the integral is this:

$$\begin{aligned} I = 2\pi \int d\theta \sin\theta[\cos\varphi - mR(1 - \cos\theta)\sin\varphi]; \\ \varphi = nt - mR(1 + \cos\theta), \end{aligned} \tag{66}$$

which is easy to evaluate, since on putting z for $(1 + \cos\theta)$ and $-dz$ for $\sin\theta d\theta$ it becomes

$$I = -2\pi \int_2^0 dz[\cos(nt - mRz) - mR(2 - z)\sin(nt - mRz)], \tag{67}$$

which can be integrated almost by inspection (the last term through integration-by-parts) and yields the desired value:

$$I = 4\pi \cos nt, \tag{68}$$

and Kirchhoff's theorem comes triumphantly through the test.

It happens however that these conditions, to which we have applied the theorem with so easy a success, are such that the result is of no value whatever in testing the undulatory theory of light.* As I have said already, the eye and the light-recording instruments register amplitude only, not phase. In the train of plane parallel waves described by (61), the amplitude is everywhere the same, and the instruments must report an impression uniformly intense. Nothing could be learned from them about the frequency or the wave-length of the light, and indeed they could not even show that there is anything periodic in the beam. The situation is no better in such a train of spherical diverging waves as (55) describes. Here the amplitude varies inversely with the distance from the centre of divergence, and the eye must report a gradual smooth decline of intensity as it moves away from that centre. In neither observation is there anything to reveal a periodicity or a wave-length, nor anything to forbid the supposition that a beam of light is a stream of straight-flying particles. What we require is a situation in which the use of Kirchhoff's theorem leads to a peculiar and striking variation of amplitude from place to place in the radiation-field—an amplitude-pattern or vibration-pattern, so to speak, depending in detail upon the wave-length of the waves, and so distinctive that if such a pattern of intensity were actually to be found in a field of light one could not but regard it as forceful evidence for the undulatory theory.

Situations which answer this requirement occur when a broad beam of waves is intercepted by a screen pierced with small apertures. In the space beyond the apertures there is a wave-motion of which the amplitude varies from place to place in a remarkable way, depending in detail upon the wave-length. When light falls onto a screen pierced with small holes, the intensity beyond the holes varies remarkably in space. The variations which are observed agree with those which are predicted from the wave-theory, when the proper value of wave-length is chosen; and this is the method of measuring wave-length. Also it is the method of measuring frequency; for the frequency of light cannot be measured directly; it must be computed by dividing the wave-

* In the actual theory of light there are several distinct quantities s—components of electric and magnetic field strengths—each of which separately conforms to equation (6), and which are interconnected in ways which need not yet concern us.

length into the speed of light. Now all the contemporary theories of the atom and of light are based upon values given for frequencies of radiation; and therefore all of them are founded on the wave-theory of light.

We will consequently next consider the propagation of waves beyond apertures, or what Rayleigh called the "effects dependent upon the limitation of a beam of light."

Propagation of a Wave-Motion beyond Apertures

Suppose that the entire plane $x = 0$ is occupied by a wall acting as a total stop to all the waves which come up against it, except where it is pierced by one or more openings; and for simplicity suppose that the oncoming waves compose a plane parallel monochromatic train,

$$s = \cos(nt - mx), \tag{69}$$

advancing from the side of negative x. The primary question is: what goes on in the region to the positive side of the screen?

The question shall be answered by approximations.

The *first approximation* consists in assuming that the wave-fronts come unaltered up to the screen, and each segment which coincides with an aperture goes straight through it and indefinitely onward, travelling unchanged in a straight line even though the surrounding portion of the wave-front has been blocked. If this were perfectly valid, there would be rectilinear propagation of light; the laws of geometrical optics would always be exact; and there would be no need for any but a corpuscular theory of radiation. Because this approximation is deficient, the wave-theory is required. Yet it is close enough to the truth to seem exact to all but the most careful observation, if the apertures are as wide as windows or even as keyholes.

The *second approximation* consists in assuming that the wave-fronts come unaltered up to the screen, and each segment which collides with a portion of the wall is swallowed up and blotted out of existence, but the wave-motion within each aperture is precisely the same as it would be if the entire wave-front passed intact across the plane $x = 0$. That is to say: the displacement on the front face of the screen is supposed to be given thus:

$$\begin{aligned} &s = \cos nt,\ \partial s/\partial x = m \sin nt \quad \text{wherever there is an aperture,} \\ &s = \partial s/\partial x \qquad\quad = 0 \qquad\quad \text{wherever there is obstruction.} \end{aligned} \tag{70}$$

This is the assumption on which are founded the conventional theories of the passage of light through a hole, or a slit, or a pair of slits, or a

diffraction-grating, or an echelon, or past a straight-edge or a solid disc or any small obstacle. Having made it, one proceeds to determine the value of s at any point beyond the screen by integrating Kirchoff's integrand all over the apertures—all over the vacant places of the screen, as if in those places only the wave-front were intact, and elsewhere it were abolished; as if the wave-front were cut into the pattern of the apertures as by a template, and each of the segments thenceforth propagated according to the law of wave-motion.

This method is fairly easy to apply, at least when the contours of the openings are simple geometrical figures, circles or rectangles for instance. In practice it is used almost always; and its results are ratified by experience. Yet it is not quite accurate.[4] I am not referring here to the ever-present possibility that boundary-conditions chosen for their mathematical simplicity may not properly describe the actual conditions in the physical world. I am referring to a mathematical, that is to say, a logical, difficulty which is inevitably fatal. A function which is equal to cos $(nt - mx)$ over arbitrary patches of the plane $x = 0$, but is always equal to zero over the remainder of the plane—such a function does not conform to the wave-equation (6). If in an actual case of light passing through a hole in a screen the phenomena conform everywhere to the wave-equation, then the boundary-condition (70) cannot be valid; if on the other hand the state of affairs in the apertures is rightly described by (70), then the wave-equation cannot be valid and Kirchhoff's theorem cannot be applied.

There is no way out of this dilemma. One must either accept the foregoing assumption frankly as an approximation, or else undertake the vastly more difficult problem of solving the wave-equation itself with the boundary-condition $s = 0$ (or some other which is deemed appropriate) for all the opaque area of the screen, and with other boundary-conditions at infinity to settle the direction from which the light is supposed to come. By this method one makes no assumption about the values of s in the apertures; they are part of the solution. But the general problem is formidable, and no less eminent a man than Sommerfeld was required for the solving of even the simplest conceivable case. Later I will mention his solution of the case in which the barrier covers all that part of the plane $x = 0$ which lies to one side of a straight line, the y-axis for instance, and the remainder of the plane is void. Meanwhile, since on the whole the second approximation is a close one, I will adopt it to explore these effects of *diffraction* which first invited the wave-theory of light, as they now are inviting that of matter; which serve to determine wave-lengths of light, and of matter;

[4] In some of the texts on optics this is not made sufficiently clear.

which set the limits for the powers of telescope and microscope, and perhaps for perception altogether; and which are responsible for the haloes and the parhelia of the sky.

Imagine then that the waves which come up to the screen from behind are plane-parallel and monochromatic, and travel in the positive sense of the x-direction, the screen itself occupying the entire plane $x = 0$. In making the "second approximation" aforesaid, we are to regard this plane as a surface where s and its gradient are zero everywhere save over certain patches—to wit, the apertures—and over these are given by the expressions:

$$\begin{aligned} s &= \cos (nt - mx)_{x=0} \quad = \cos nt, \\ \partial s/\partial x &= m \sin (nt - mx)_{x=0} = m \sin nt. \end{aligned} \tag{71}$$

We are then to determine the value s_0 of s at any field-point P anywhere before the screen—anywhere in the region $x > 0$—by forming Kirchhoff's integral over these apertures:

$$4\pi s_0 = \int dS \left[\cos (n, r) \frac{\partial}{\partial r} \frac{U}{r} - \frac{1}{r} \frac{\partial U}{\partial n} \right]. \tag{72}$$

Over the rest of the plane $x = 0$ the integrand vanishes. Since, however, Kirchhoff's theorem involves an integration over an entire closed surface surrounding P, we ought in strictness to extend the integral over some far-flung surface completing the enclosure; as for instance a hemisphere seated upon the plane $x = 0$, sufficiently great in radius to contain P and all the apertures. This is always neglected, possibly because in practice the wave-motion over such a surface would as a rule be too chaotic to produce any regular effect at P.[5]

In the integrand of (72), r stands for the distance from P to any area-element dS of an aperture; the positive r-direction is measured from P through dS in the direction from front to back; the positive n-direction is the forward-pointing normal to dS, and therefore is identical with the positive x-direction. Remembering the definitions of U and its derivatives, one easily sees that:

$$\begin{aligned} U &= \cos (nt - mr); \\ \partial U/\partial r &= \partial U/\partial x = \partial U/\partial n = m \sin (nt - mr). \end{aligned} \tag{73}$$

It will be convenient to give the symbol θ to the angle between the posi-

[5] Certainly it cannot be argued that the effect from a distant surface is necessarily too small to be noticed at P; we have just seen that in a field of plane-parallel waves it is the same for any spherical surface, no matter how great the radius.

tive x-direction and the line from dS to P, so that $\cos(n, r) = -\cos\theta$. Consequently:

$$4\pi s_0 = \int dS\left[-\frac{1}{r^2}\cos(nt - mr) - \frac{m}{r}(1 + \cos\theta)\sin(nt - mr)\right]. \quad (74)$$

One is tempted to say that the quantity under the integral sign is the contribution made by the element-of-wave-front dS to the value of s at P. This notion facilitates both thought and description, and I will adopt it, but with a warning. The danger is that one may come to think of an element-of-wave-front as an independent entity, capable of existing by itself in the medium regardless of what other elements-of-wave-front adjoin it or stand elsewhere. This is unpermissible, for the same reason which makes the method that I am now expounding an approximate and not a rigorous one. Were one of these elements of wave-front alone in the medium, the function s would not conform to the wave-equation. Therefore if we call the expression

$$ds_0 = \frac{1}{4\pi}dS\left[-\frac{1}{r^2}\cos(nt - mr) - \frac{m}{r}(1 + \cos\theta)\sin(nt - mr)\right] \quad (75)$$

the *contribution* of the element-of-wave-front dS, we must always remember that it cannot be isolated, but—like the donations to certain endowments—is given only under the condition that other elements also contribute.[6]

The contribution of dS, then, is made up of two terms, one varying inversely as r^2 and the other inversely as r/m. At great distances the latter must increasingly outweigh the former; and "great distances" in this context signify those which are much greater than $1/m$—that is to say, very many times as great as the wave-length of the light. Now as the wave-lengths of most kinds of light are less than .001 mm., a field-point where observations can actually be made must necessarily be distant by many wave-lengths from the screen. Hence it is customary to ignore the first term in the expression (75) and in its integral, and write for the contribution of dS:

$$ds_0 = -\frac{1}{4\pi}dS\frac{m}{r}(1 + \cos\theta)\sin(nt - mr). \quad (76)$$

This expression is the approximate description of what in an earlier

[6] The reader may notice that whereas in dealing with a closed surface surrounding the field-point the n-direction was defined as that of the "inward-pointing normal," there is no way of discriminating between the two senses of the normal to an isolated area-element. This causes an ambiguity in the sign of the contribution; for reversing the sense of the normal reverses the signs both of $\partial U/\partial n$ and of $\cos(n, r)$. The ambiguity is always, I think, physically trivial.

passage I called the "influence" or the "wavelet" which spreads out from the element-of-wave-front in all directions.

Examining it factor by factor, one sees:

(*a*) that the amplitude of the wavelet varies inversely as the distance r from the starting-point, which seems natural;

(*b*) that the wavelet is not isotropic, its amplitude diminishing according to the law $(1 + \cos \theta)$ from a maximum value in the forward to zero in the rearward direction. This is commonly stated as the reason why waves can be propagated in one direction only, not necessarily both forward and backward at the same time;

(*c*) that for waves of the same amplitude and different wave-lengths the amplitudes of the wavelets stand in the inverse ratio of the wave-lengths—the shorter the waves, the more powerfully they are diffracted;

(*d*) that the wavelet from any point is constantly one quarter of a cycle in advance of the primary wave, varying as $-\sin nt$ whereas the wave varies as $\cos nt$.

The advance-in-phase and the factor m in the amplitude enter, it is clear, because the "wavelet" represents the second term in (75)—the term which involves the slope $\partial s/\partial x$ of the wave-function, not the wave-function itself. One might say that the cyclic variation of $\partial s/\partial x$ stirs up a relatively far-reaching commotion in the medium, while the disturbance which the cyclic variation of s excites is rapidly attenuated and mostly negligible. Formerly the factor m and the advance-in-phase seemed unnatural and very strange; for they antedated the theorem of Kirchhoff by sixty years, having been forced upon Fresnel before 1820—and this invites an allusion to history.

Though it is in connection with Huyghens' principle that one commonly hears of wavelets, that principle itself amounts to a denial of nearly every quality which we associate with the ideas of wavelet or wave. Not only are the "wavelets" of Huyghens' construction quite devoid of anything undulatory or periodic; the construction itself is based on the assumption that there is only one point on each where the amplitude is appreciable—the point on the prolongation of the normal from the primary wave-front (corresponding in my notation to $\theta = 0$). But to say that a disturbance is transmitted by wavelets such as these is to say that it is transmitted in concentrated form along lines or rays—which is the same thing as saying that it travels like corpuscles. Huyghens' principle in fact leads straight to the doctrine of the rectilinear propagation of light, and fails either to predict or to explain the phenomena which require a wave-theory.* The accredited

* I am not prepared to say that this is true of the applications to crystal optics.

founder of the wave-theory of light invented in reality a novel language for expressing the corpuscular theory!

Fresnel however invested these wavelets of Huyghens with some of the properties which entitle them to the name. He supposed that the amplitude was distributed widely over each, not confined to the point $\theta = 0$, though greatest at that point; he thought that it diminished slowly with increase of θ, though he did not suggest the precise factor $(1 + \cos \theta)$ nor any other; and he thought that it varied inversely as distance. Further, he endowed it with a periodicity. Thus far, he was right. But naturally he supposed that the cause of the wavelet was the cyclic variation of the wave-function s, and therefore he presumed that it started out in consonance of phase with the primary wave; and he did not insert the factor m. However when he came to test his ideas in somewhat the same way as Kirchhoff's theorem has been tested in these pages—by applying them to a case where the required result was known *a priori*—he was unable to derive the proper answer, except by introducing the factor m and the advance-in-phase; and thenceforth they have figured in the theory of diffraction, indispensable and until the day of Kirchhoff inexplicable.

To return to the problem of determining the wave-motion beyond the apertures: under the approximations stated, it is mathematically quite definite. The solution is the value of the integral:

$$s_0 = -\frac{1}{4\pi}\int dS\left[\frac{m}{r}(1 + \cos\theta)\sin(nt - mr)\right], \qquad (77)$$

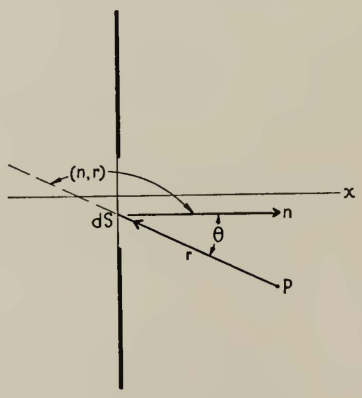


FIG. 2

extended over the apertures; r standing for the length of the line joining the field-point P with the element-of-wave-front dS, and θ for the angle between this line and the perpendicular dropped from P to the plane of the screen.

A simple, instructive, and historically famous example is that of the circular hole.

Diffraction from a Circular Aperture

If the propagation of waves were rectilinear, their amplitude would be constant along every line passing normally across an aperture. Such however is as far as possible from being the truth, as we can easily learn by evaluating the wave-motion along the "axis" of a circular hole—that is, the line passing through the centre of the hole perpendicular to the plane of the screen. Locate the centre of the circle at the origin, so that its axis is the axis of x. Denote by R the radius of the circle, by x_0 the coordinate of the field-point P located anywhere upon the axis. All points on any circle centred at the origin being equidistant from P, we may divide the area of the hole by concentric circles into annular elements-of-area. Denote the radius of such a one by p, its breadth by dp; then for it:

$$dS = 2\pi p dp; \qquad r^2 = x_0^2 + p^2; \qquad \cos\theta = x_0/r, \tag{78}$$

and the limits of integration are $p = 0$ and $p = R$.

The problem is now stated in full; but it is very much simplified if the distance x_0 from screen to field-point is very many times as great as the width R of the aperture; and this in practice is commonly the case. Then to first approximation,

$$r = x_0, \qquad \cos\theta = 1, \tag{79}$$

and these values are close enough to the correct ones to suffice for the multipliers of the sine-function in (77); but in the argument of the sine, r is multiplied by m, and a variation of only half a wave-length in r entails a complete reversal of the function; hence in the argument we must proceed to second approximation, and write

$$mr = mx_0 + mp^2/2x_0. \tag{80}$$

Making these substitutions in (77), we have finally:

$$s_0 = -(m/x_0)\int_0^R p \sin(nt - mx - mp^2/2x_0)dp$$

$$= \cos(nt - mx_0) - \cos(nt - mx_0 - mR^2/2x_0) \tag{81 a}$$

$$= \sqrt{2(1 + \cos mR^2/2x_0)}\cos(nt - mx_0 - a). \tag{81 b}$$

Interpreted, these equations tell the startling fact that along the axis of the hole the amplitude, far from being constant, varies in a

gradual and cyclic way between zero as one extreme and double the amplitude of the unintercepted wave-train as the other. As the field-point is displaced along the axis towards or away from the aperture, as the aperture itself is expanded or contracted, doubled agitation succeeds upon quiescence and quiescence upon agitation; and the opening, far from serving as a window to let a segment of the oncoming wave-train pass unaltered by, acts as an agency for producing a curious pattern of varying amplitudes over the region before it.

Now these are precisely the conditions under which, as I remarked before, one can arrange a test of the wave-theory of sound or light; for here we have the amplitude varying from point to point, in a pattern depending in detail upon the wave-length. Experience of light reveals just such a pattern; when parallel light is shed normally upon a screen pierced with a small and accurately rounded hole, the illumination in the axis of the hole passes alternately through maxima and minima as the observer recedes along it. Fresnel was led in a curious way to discover the minima. The French Academy having offered a competitive prize for a study of diffraction—an action instigated, it appears, by adherents of the corpuscular theory of light, who expected that a thorough knowledge of the phenomena of diffraction would demolish the support which they were vaguely supposed to provide for the wave-theory—Fresnel conducted a research and submitted a memoir which ranks among the classics of physical science. It went for judgment to an illustrious committee of five,[7] one of whom, the very eminent mathematician and physicist Poisson—who had been an upholder of the corpuscular theory—promptly deduced the law of the maxima and minima along the axis from Fresnel's conception of the wavelets. He imparted this prediction to the author of the memoir; and in a note appended to the published version, Fresnel has left it on record that he looked for a minimum and found it "like an inkspot" in the centre of the field before the hole.

Equation (81) shows further that the amplitude at any point upon the axis must vary to and fro between the same two extremes—zero, and double the amplitude of the unhindered waves—as the hole expands or shrinks. Wood has described how this may be observed with an iris diaphragm. For an observer stationed at a fixed point upon the axis at a distance x_0 from the hole, the amplitude falls to zero whenever the radius of the circle has one of the values determined by the condition

$$mR^2/2x_0 = \text{even integer multiple of } \pi, \tag{82}$$

[7] Arago, Biot, Gay-Lussac, Laplace and Poisson. It would be hard to assemble a more distinguished group at any time or place.

that is to say, whenever

$$x_0 = mR^2/2k\pi, \qquad k = 0, 2, 4, 6, \ldots, \tag{83}$$

and attains its maximum value, double the amplitude of the uninterrupted waves, whenever

$$x_0 = mR^2/2k\pi, \qquad k = 1, 3, 5, 7, \ldots. \tag{84}$$

Imagine circles drawn upon the plane of the screen, with their common centre at the origin and their radii R_1, R_2, R_3, . . . prescribed by the equations,

$$mR_k^2/2\pi x_0 = k, \qquad k = 0, 1, 2, 3, 4, \ldots. \tag{85}$$

They divide up the plane of the screen into a tiny central circular area and a series of surrounding rings. These are the "Fresnel zones" relative to the point x_0 where the observer is placed. If the circular hole comprises an odd number of the zones, the wave-motion at x_0 attains its maximum; if an even number, the wave-motion vanishes—there is silence or darkness. It seems as if the first, third, fifth and other odd-numbered zones brought light, and the second, fourth and other even-numbered zones destroyed it.

It is equally easy to find the wave-motion along the axis of an annular opening—that is to say, a circular hole partly filled by a concentric circular stop. Denote by R_0 the radius of the stop and by R the radius of the hole; then the limits of integration in (81) are superseded by $p = R_0$ and $p = R$, and the amplitude along the axis varies thus:

$$A = \sqrt{2[1 - \cos m(R^2 - R_0^2)2x_0]}. \tag{86}$$

This contains the surprising conclusion that the maxima of amplitude along the axis are as great as they would be if the stop were removed, though they may be differently placed. An observer properly stationed should see the light brighten when the obstacle is inserted; it may even be brighter than when the obstacle within the hole and the wall surrounding it are totally removed, leaving no hindrance to the onward march of the waves.

The conclusion still holds good when the boundaries of the circular hole retire to infinity, leaving nothing but an opaque disc in an otherwise uninterrupted stream of plane parallel waves; although the approximations made in the foregoing pages are then no longer valid, and equation (86) is not to be employed. Experience however shows that when a small and accurately rounded circular disc is immersed in a beam of parallel light there is a bright spot—more precisely speaking,

a bright core—along the axis of the geometrical shadow. Poisson forecast this also when Fresnel's memoir came before him, and seems to have thought that it would make an *experimentum crucis*, for another member of the committee—Arago—has recorded that he tested the prediction when Poisson made it. He found the bright spot in the centre of the shadow of a circular disc. It is said that Delisle had found and recorded it already, but the record had slipped into oblivion.[8]

We take up now the problem of determining the wave-motion away from the axis—otherwise expressed, that of determining the distribution-of-amplitude over any plane parallel to the plane by the screen.

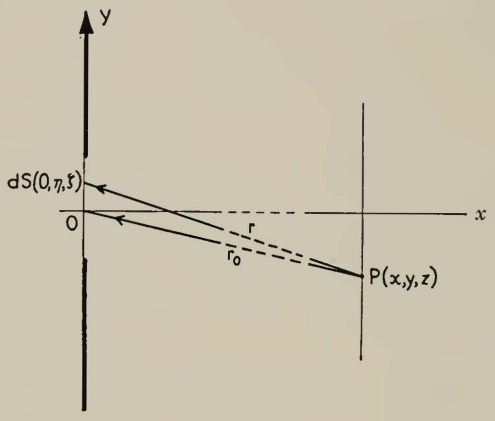


FIG. 3

Denote by (x, y, z) the coordinates of any field-point and by $(0, \eta, \zeta)$ those of any area-element dS of the aperture; by r, as heretofore, the distance from P to dS, and by r_0 the distance from P to the origin. Then

$$r^2 = x^2 + (y - \eta)^2 + (z - \zeta)^2 = r_0^2 - 2y\eta - 2z\zeta + \eta^2 + \xi^2. \quad (87)$$

As heretofore r and r_0 shall be supposed to be very many times as great as the dimensions of the apertures, and therefore as the greatest values attained by η and ζ; therefore, to first approximation,

$$r = r_0, \qquad \cos\theta = x/r_0, \quad (88)$$

[8] It is interesting to notice why an accurately circular disc is required to show the bright spot in its best development. Take the case of the aperture, since we already have its suitable equation (81). A nearly but not quite circular hole may be regarded as made up of sectors, each with a different radius. For each of these the upper limit of the integral in (81) would be different, and therefore the condition (84) for doubled amplitude could not be realized for all at once. There would be a wave-motion along the axis, but not the regular alternation of maxima and minima nor the sharply outstanding brightness at the maxima.

and to second approximation,

$$r = r_0 - (y\eta + z\zeta)/r_0. \tag{89}$$

Into equation (77) we insert the first-approximation values of r and $\cos\theta$ in the multipliers of the sine-function, but the second-approximation value of r into the argument of the sine. Therefore we have, for the value of s at the very distant point (x, y, z), the expression:

$$s = -\frac{1}{4\pi}\frac{m}{r_0}\left(1 + \frac{x}{r_0}\right)\int\int d\eta d\zeta \sin\left(nt - mr_0 - m\overline{y\eta + z\zeta}/r_0\right). \tag{91}$$

It is expedient to introduce the three direction cosines of the line extending from the origin to the field-point, the cosines of the angles between it and the coordinate axes:

$$\alpha = \cos(x, r_0) = = x/r_0; \qquad \beta = y/r_0; \qquad \gamma = z/r_0. \tag{92}$$

Then, with a slight additional transformation, we convert equation (89) into:

$$\begin{aligned} s = \text{const.}\,(1+\alpha)\Big[&\sin(nt - mr_0)\int\int d\eta d\zeta \cos m(\beta\eta + \gamma\zeta) \\ &- \cos(nt - mr_0)\int\int d\eta d\zeta \sin m(\beta\eta + \gamma\zeta)\Big] \\ = \text{const.}\,(1+\alpha)&[C\sin(nt - mr_0) - S\cos(nt - mr_0)], \end{aligned} \tag{93}$$

the symbols C and S being traditional for these integrals.

The coordinates of the field-point have disappeared, leaving only the cosines which define its direction as seen from the origin. This means that we have here the formula for the wave-motion over any plane parallel to the screen and infinitely far away, in terms of the directions in which its various points are seen. The words "infinitely far away" sound formidable; but it is not necessary to depart for infinity, in order to find a plane where (93) describes the state of affairs. There is an artifice for bringing the infinitely distant plane up to a convenient nearness; an artifice known as a *lens*. When a converging lens is set up before the apertures, the wave-motion predicted by the formula (93) for all points infinitely far away upon the line with direction cosines (α, β, γ)—this wave-motion occurs at the point where the line intersects the focal plane of the lens. Therefore we may regard equation (93) as the description, according to the wave-theory of light, of the distribution-of-amplitude in the focal plane of the lens. (To convert the cosines into coordinates in that plane, it is sufficient to multiply each by the focal length of the lens.)

Returning now to (93), it is evident that the problem is solved when the integrals are evaluated; in particular the amplitude is given by the formula,

$$A = \text{const.}\ (1 + \alpha)\ \sqrt{C^2 + S^2}. \tag{94}$$

Whatever the shape of the aperture or apertures, the values of the integrals can be determined as closely as may be desired; and in two instances which happily are the most frequent and useful—those of the circular and the rectangular openings—the integrations lead directly to familiar functions.

Diffraction Patterns in the Focal Plane of a Lens

If the origin is located at the centre of the circle, the integral S vanishes—for the value of the sine-function contributed by each area-element is annulled by the value contributed by the element symmetrically placed to the other side of the centre—and the integral C for the same reason becomes this:

$$C = \int\int \cos(m\beta\eta) \cos(m\gamma\zeta) d\eta d\zeta. \tag{95}$$

By putting $\gamma = 0$ and then integrating, we shall obtain the distribution of amplitude along the line passing through the centre of the diffraction-pattern and parallel to the axis of y; but this, by reason of the circular symmetry of the entire system, is the same as the distribution of amplitude along any radius passing through the centre of the diffraction-pattern, and therefore is all we need. In the expression so obtained, replace the Cartesian coordinates heretofore used in the plane of the screen by polar coordinates ρ and φ; then we have

$$C = \int\int \rho \cos(m\beta\rho \cos\varphi) d\rho d\varphi, \tag{96}$$

the limits of integration being 0 and R (the radius of the aperture) for ρ, and 0 and 2π for φ.

The integral C is proportional to the Bessel function of order unity of the variable $mR\beta$:

$$C = 2(\pi R/m\beta) J_1(mR\beta). \tag{97}$$

This is a function which like the sine vanishes at intervals, though not at equal intervals. The centre of the diffraction-pattern is therefore encircled by concentric rings over each of which the wave-motion vanishes; between each pair of these there is a zone where the amplitude differs from zero and varies, attaining a maximum somewhere near the middle of the zone. In the focal plane of the lens there are ring-shaped zones of light, surrounded and divided by dark circles; these are the "fringes."

This system of annular fringes is the *image* produced by a lens on which plane-parallel light falls normally through a circular aperture; also when there is no screen before the lens, for being circular it serves as its own aperture. Now plane-parallel light is such as originates in an infinitely distant luminous point, or—what comes to the same thing —any luminous object so distant that neither the curvatures of the wave-fronts proceeding from its various parts nor the angles between the directions in which these lie are appreciably large; a star, for instance. The image of a star in the focal plane of a telescope objective is therefore not a point, however far away the star may be; it is a system of rings. So it is in the eye, the pupil serving as the aperture; but the inner rings are in both cases so narrow and the outer rings so faint that they appear condensed into a point. Magnification of the fringes in the telescope by the eyepiece brings them into view, and so they set a limit to the value of magnification; for it is of no avail to be able to examine an image more minutely if all that can be examined is the consequence of the disturbance produced in the incoming waves by the finiteness of the lens.

The limitation which the law of propagation of light thus sets upon the formation of images is very important. The simplest possible illustration is furnished by a double star. Let the telescope be directed upon such a pair of stars so that the light from one component falls normally upon the lens, the light from the other component at any angle of which I denote the complement by φ; thus φ stands for the angular distance between the two stars in the sky. Now I have not hitherto treated the case of light falling otherwise than normally upon the screen containing the apertures, which in this case is nothing but the plane of the objective. The extension however is immediate. Orienting the y-axis in the plane of the screen so that the direction of propagation of the waves coming from the stars lies in the xy-plane, we have for the wave-function in the region extending up to the aperture from behind:

$$s = \cos (nt - mx \cos \varphi - my \sin \varphi), \tag{98}$$

and therefore in the plane of the aperture ($x = 0$) we have, instead of the values given in (71), these:

$$\begin{aligned} s &= \cos (nt - my \sin \varphi), \\ \partial s/\partial x &= m \cos \varphi \sin (nt - my \sin \varphi), \end{aligned} \tag{99}$$

and for the value of s at any point in front of the aperture we have, instead of (77), this value:

$$s_0 = -\frac{1}{4\pi}\int dS\left[\frac{m}{r}(\cos\theta + \cos\varphi)\sin(nt - my\sin\varphi - mr)\right],$$

and there are corresponding changes in the values of the integrals C and S which determine the amplitude. To first approximation—that is to say, when φ is not too great—the result is, that the diffraction pattern of one star is like that of the other, but shifted sidewise. The angular displacement between the centres of the two fringe-systems is the same as the angular displacement between the two stars. The question now arises: how far apart must the two fringe-centres be, that the two families of rays may be securely told apart?

Such a question of course cannot be definitely answered; the answer would depend upon the acumen and the experience of the observer. The conventional response is, that the two systems of rings are surely distinguishable if the centre of one lies upon the first dark circle of the other. Now the angular radius of the first dark ring, i.e., the value of β for which the Bessel function of (97) first vanishes, is 1.22 times the ratio of the wave-length of the light to the diameter of the aperture; for green light in the largest available refracting telescope this amounts to about an eighth of a second of arc. This then is nearly the least angular separation between two stars which are distinguishable; a pair or a group much closer together would appear as one, not through any avoidable defect of the telescope nor through any insufficiency of the eyepiece but through the laws of propagation of light themselves, working to prevent the formation of an image indefinitely sharp.

For a rectangular aperture the integrals C and S are extremely easy to evaluate. The diffraction-pattern is a criss-cross of dark lines, intersecting at right angles and bounding rectangular areas of light, similar in shape to the aperture but oriented at right angles to it. If the rectangle is prolonged indefinitely and so becomes an infinitely long slit, the diffraction-pattern becomes a sequence of parallel bands separated by dark lines normal to the length of the slit. If then a multitude of identical slits are cut into the screen at equal intervals side by side, a new periodicity is superposed upon the periodicity of the waves, and out of the interaction of these two there come diffraction-patterns much more sharp and striking than any which a single aperture, however shaped, is able to produce. These will be considered in the following chapter.

Recent Developments in the Process of Manufacturing Lead-Covered Telephone Cable [1]

By C. D. HART

THE manufacture of telephone cable consists essentially of insulating copper wire with paper, twisting two insulated wires together to form a pair, again twisting to form a quad if quadded cable is to be made, stranding these pairs or quads into a compact core, removing moisture, covering the core with a continuous sheath of lead or lead alloy, testing the completed cable and packing it for shipment.

In order to bring out clearly some of the recent developments in manufacturing processes it is necessary to review the beginning of the art.

The idea of using cables for telephonic communication goes back to about 1878. In a talk given in London by Dr. Alexander Graham Bell he stated "It is conceivable that cables of telephone wires could be laid underground, or suspended overhead, communicating by branch wires with private dwellings, country houses, shops, manufactories, etc., uniting them all through the main cable with a central office where the wires could be connected as desired, establishing direct communication between any two places in the city."

About two years later, or in 1880, the idea became a fact and wires enclosed in sheath were used across the Brooklyn Bridge.

The insulation used on these early cables was gutta-percha or rubber but these materials were not very satisfactory for land telephone cables. A little later sisal and cotton were used and the cable core was impregnated to prevent the entrance of moisture and then drawn into successive lengths of lead pipe previously extruded and laid out in straight pieces, the different lengths being then joined together by means of plumber's joints. Impregnation was resorted to because it was difficult to obtain a lead sheath which was entirely free from defects.

By about 1890 paper ribbon had been introduced as a substitute for cotton and similar insulations, effecting, of course, a great saving in space and therefore in sheathing material and cost.

Fig. 1 shows a group of insulating machines used about 1892. With these machines paper ribbon was wound from a spool mounted eccentrically with the wire and the insulating speed was necessarily very slow.

[1] Presented at the Regional Meeting of District No. 5 of the A. I. E. E., Chicago, Ill., Nov. 28–30, 1927.

Fig. 2 shows the twisting machines used at that time for twisting pairs. These machines were crude and operated at a low speed.



Fig. 1—Old insulating department about 1892



Fig. 2—Old twisting department about 1890

Fig. 3 is of an old stranding machine consisting of one drum only as the cable cores were built up one layer at a time and the core was run

through the stranding machine as many times as there were layers in the finished core.

The old process of pulling cable core into lead pipe is illustrated in Fig. 4. This picture was posed a few years ago, and the man standing in the foreground was one of a gang who formerly did this work.



Fig. 3—Old stranding department about 1890

A forward step in design of insulating equipment was made with the use of pads concentric with the wire which permitted very much higher insulating speeds and very much reduced paper breakage. The twisting machines were also modified to reduce uneven twisting and permit greater speed.

Another step was the development of multiple drum stranders permitting a number of layers or complete small cables to be made in one operation. Also, extrusion presses were improved so that a continuous sheath of lead alloy could be extruded directly on to the cable core, eliminating the pulling-in operation.

During the period from about 1900 to 1920 many changes were made to increase output and improve the quality. Improvements in machines made possible the use of thinner and narrower insulating papers so that a greater number of pairs of wires could be placed within the same cross-sectional area, tending greatly to decrease the cost per circuit. Cables made about 1888 contained 50 pairs of 18-gauge conductor. By about 1902 improvements had been made which permitted 606 pairs of 22-gauge wire to be put into a sheath of $2\frac{3}{8}$ in.

inside diameter which is the maximum size of sheath which has been found generally economical in telephone plants in this country. By 1912 further improvements in manufacturing equipment made it possible to use insulating paper of even smaller dimensions and to get 909 pairs of wire into the same diameter of sheath.



Fig. 4—Pulling core into lead pipe, method used prior to 1894

On account of increased congestion in the densely populated sections of the larger cities, there was continued demand for more pairs of wire per cable, and in 1914 the first 1,212-pair 24 A.W. gauge cables were produced. This 24-gauge wire was insulated with paper $^{5}/_{16}$ in. wide and $2\frac{1}{2}$ mils thick. The mutual capacitance between the two wires of a pair in this cable averages about .079 microfarad per mile, which allows a normal margin below the guaranteed value shown in Table 1.

TABLE 1

A.W.G.	Standard Sizes—Pairs	Average A. C. Capacitance Guarantee m.f. per Mile	Principal Uses
13	11 to 76	.071	Toll entrance and long trunks.
16	11 " 152	.071	
19	6 " 455	.090	Trunk and long subscriber lines.
22	11 " 909	.089	Subscriber lines.
24	11 " 1212	.085	Short subscriber lines.

The insulation withstands a potential test of 500 volts (maximum instantaneous value). The increasing number of pairs per cable and

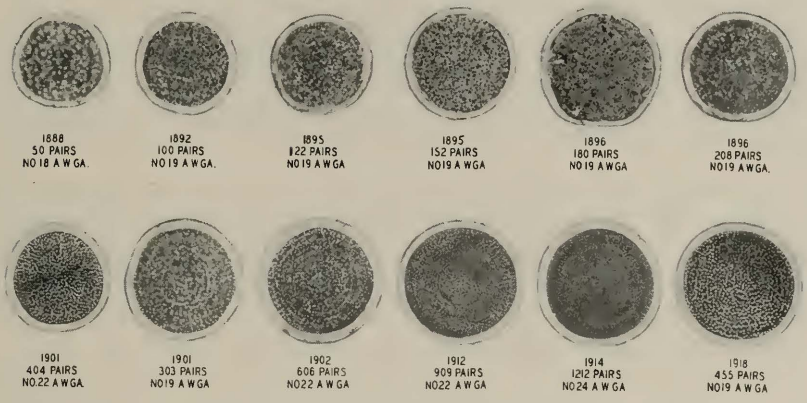


Fig. 5—Principal stages in the development of paper-insulated cable

the corresponding decreasing cost per mile of circuit resulting from the changes described is shown in Figs. 5 and 6.

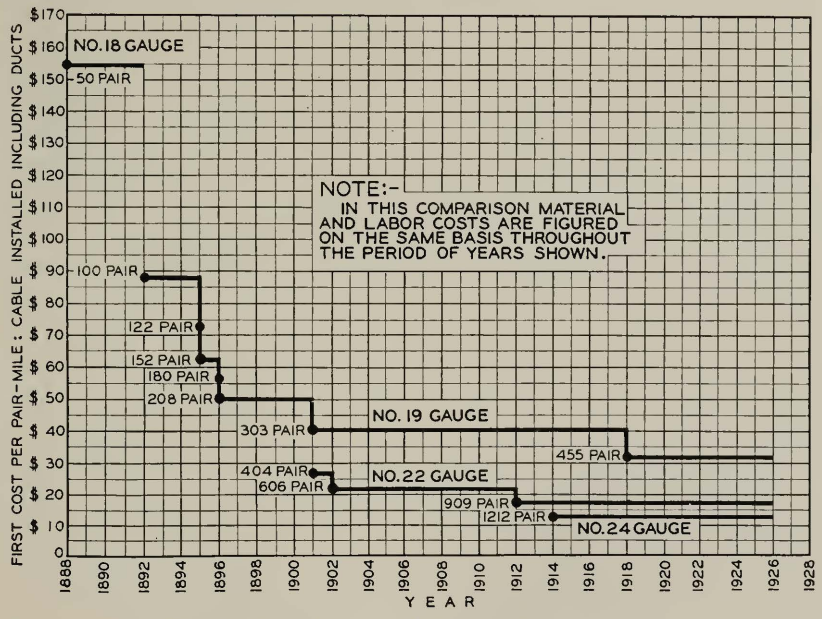


Fig. 6—Cost of a mile of circuit in full-size cable

With the growth of large office buildings and further increases in the demand for telephones in the great cities, even 1,212 pairs of wire per

cable were in some cases found to be inadequate, and in answer to the demand a cable has been developed containing 1,818 pairs of 26 A.W.G. wires within a sheath having an inside diameter of $2\frac{3}{8}$ in.

These wires are insulated with paper $\frac{7}{32}$ in. wide and $1\frac{3}{4}$ mils thick by the use of specially designed insulating heads and, instead of being stranded in reverse layers as is the case with older types of cables, they are first stranded in groups of 101 pairs, 18 of these groups being then cabled together to form a compact core.

This method of cabling, called the "unit" type to distinguish it from the layer type, has several advantages, particularly in splicing in the field. Development work on this 1,818-pair cable is not yet complete but there is no reason to doubt that, if there is a demand for a 2,400-pair cable, the demand will be met.

For convenient reference Table 1 has been shown giving the specified limiting characteristics of some of the standard types of non-quadded cables. From the table it will be seen that the larger gauge cables are used mostly for trunk work and the smaller gauges for connections to subscribers. While the electrical characteristics of these non-quadded cables are of prime importance, they do not demand quite the extreme refinement in manufacturing processes required for quadded cables.

The discussion so far has been confined mainly to cable intended for local service, that is, cable providing conductors to connect subscribers directly with the central office and different offices with one another. Gradually, the network of long lines connecting different exchange areas or cities grew and while the early lines were mostly open-wire, it was necessary to provide cable in and near the larger cities to bring these lines into the central offices. Most of the long lines were operated on the phantom principle where four wires are combined to provide two ordinary pair circuits and a third or phantom circuit which uses the four wires simultaneously. It was, therefore, necessary to provide cable for these toll entrances which could also be operated on the same phantom principle. More recently many long toll lines have been placed for their entire length in cables of this type.

One of the greatest difficulties in providing this type of cable was that of building it with sufficiently good electrical balance to avoid serious interference or "crosstalk" between the various circuits in the same four-wire group or "quad," such crosstalk being especially liable to occur because practically all of these lines are loaded. For a given degree of imperfection in capacitance balance, crosstalk is much more serious if the line is loaded than otherwise. A very considerable amount of work was necessary to determine the principles of design and manufacture which have the most influence in bringing about the best balance reasonably attainable.

The specified limiting degree of unbalance of the capacitance in quadded cable is indicated in Table 2, and Fig. 7 is a diagram showing the capacitances involved and a brief explanation of them.

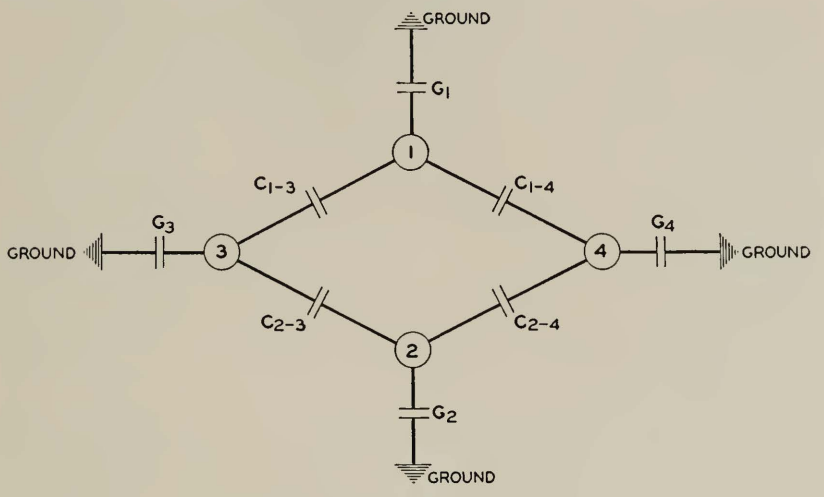


Fig. 7—Diagram showing the capacities involved in capacity unbalances between circuits

TABLE 2

Capacitance in m.f. per mile		Capacitance Unbalance in m.m.f. per 500 ft. length					
Pair	Quad	Side to Side		Phantom to Side		Phantom to Phantom	
Av.	Av.	Av.	Max.	Av.	Max.	Av.	Max.
.068	.112	30	100	120	200	60	600

[1] Class I Unbalances—Phantom to Side

1, 2, 3 and 4 represent the four wires of a quad, of which 1 and 2 form one pair and 3 and 4 form the other pair.

Unbalance between Phantom and Side $1\text{–}2 = 2[C_{1-3}+C_{1-4}-(C_{2-3}+C_{2-4})]+G_1-G_2$
Unbalance between Phantom and Side $3\text{–}4 = 2[C_{1-3}+C_{2-3}-(C_{1-4}+C_{2-4})]+G_3-G_4$

Class II Unbalances—Side to Side

1, 2, 3 and 4 represent the same as in Class I Unbalances.

Unbalance between Side 1–2 and Side $3\text{–}4 = C_{1-4}+C_{2-3}-(C_{1-3}+C_{2-4})$

[1] Capacitance Unbalances involve differences of Direct Capacitances. See G. A. Campbell, *Bell System Technical Journal*, July 1922.

CLASS III UNBALANCES—BETWEEN CIRCUITS IN DIFFERENT QUADS

Unbalances between two phantoms, or between pairs not in same quad, or between a phantom and a pair not in same quad, in each case $= C_{1-4}+C_{2-3}-(C_{1-3}+C_{2-4})$ in which, for

(*a*) *Phantom to Phantom*, 1 represents the two wires connected in parallel to form one pair of a quad, 2 represents the two wires of the quad, and 3 and 4 represent similarly the pairs of another quad.
(*b*) *Pair to Pair*, 1 and 2 represent the two wires of a pair and 3 and 4 the two wires of another pair not in the same quad.
(*c*) *Phantom to Pair*, 1 and 2 represent a phantom as in (*a*) and 3 and 4 a pair as in (*b*).

The type of quad now most commonly used in toll cables in this country is known as the multiple twin type and consists when completed of two twisted pairs which are again twisted around each other. Differently colored wrappings of cotton around the several pairs hold the two wires of the pair together and afford means of identifying various types of quad and pair as used, for example, in the segregation of the circuits operating in different directions in the so-called four-wire circuits.

A type of quad construction different from that described above and commonly known as the "spiral four" type of quad has been used more extensively abroad than here. In this construction four wires are twisted together in such a way that at every position each wire occupies approximately a corner of a square and the two diagonally opposite conductors are used to form a pair.

This construction has the merit of very low mutual capacitance of the pairs, but the disadvantage of very high mutual capacitance of the phantom. It has also been found more difficult with this construction to obtain sufficiently good balance to give satisfactory loaded phantom circuits. This type of quad has, therefore, in some cases been used without utilizing the phantom circuits. The loss of these phantom circuits is less than it might seem at first sight because, on account of the inherently lower pair capacitance for a given space per pair, more wires can be placed in the same space for a given capacitance than with other types of construction.

Another characteristic which under certain conditions is important is the alternating current conductance or leakance. The leakance which is measured in micromhos is that property which determines, under given conditions of potential and frequency, the losses in the insulation. These losses become of greater importance when the cable is loaded than when non-loaded and also of relatively greater importance when the conductors are large because then the dielectric losses become relatively greater in comparison with the lower losses in the decreased resistance of the conductor. For this reason many of the

large gauge loaded toll cables are treated with a special drying process to diminish the leakance.

Under-water Cables

Either quadded or non-quadded cable may be used on occasion for crossing rivers, bays, etc., and in these cases the lead-covered cable is protected by being first served with two or three layers of jute roving impregnated with tar, then wound with galvanized steel armor wire, and again served with jute yarn, impregnated with an asphalt compound, although in many cases at present this outer serving of yarn is omitted. In case of injury causing an opening in the sheath of such a cable, water may enter the interior and interrupt the service. It is also liable to penetrate for a considerable distance and thus ruin a substantial length of cable which it then becomes necessary to replace. To diminish the amount of cable damaged in this way, this type of cable is sometimes made with a very large amount of paper insulation crowded into a small space to make the cable within the lead pipe very dense. The swelling of this paper as it becomes wet tends to retard the penetration of water and to diminish the amount of cable damaged.

This dense core construction has, however, the objection that it tends to produce circuits of lower transmission efficiency on account of the higher capacitance and leakance obtained. For this reason cables for this purpose in many cases are made with less dense core construction similar to that used in land cables but with the core treated so as to provide water barriers at frequent intervals to prevent or greatly diminish the passage of water through the barrier, commonly known as a "plug," so that the damage resulting from an injury to the sheath is substantially confined to the portion between two consecutive plugs.

The Cable Sheath

One of the outstanding developments in cable manufacture which occurred about 1911 was the substitution of 1 per cent antimony in lead cable sheath for 3 per cent tin. The use of tin alloyed with lead for cable sheath had been instituted many years before, as it had been found that such sheath was more durable than sheath composed of lead alone and had better mechanical characteristics.

Exhaustive tests showed that lead-antimony alloy sheath is equal in quality to lead-tin alloy and, although its use required the development of improved methods of mixing and extrusion, it has resulted in large cost savings.

Another decided improvement introduced later was the substitution of vacuum drying ovens for the old gas or steam-heated air ovens.

It was found that the drying time using vacuum ovens was reduced to about one third as compared with hot air ovens, and improved quality and large cost savings resulted.

Before the war the average demand for telephone cable in this country amounted to about two hundred million conductor feet per week. During and after the war this demand steadily increased until now it amounts to about six hundred million conductor feet per week or about thirty billion feet per year, requiring annually forty thousand tons of copper wire, seventy-five thousand tons of lead, and six thousand tons of insulating paper.

Cable-making Machinery

In planning for the manufacture of this quantity of cable, the design of all machinery was reviewed and changes made wherever possible to improve quality or increase output.

A great deal of work was done in improvement of insulating machines, and a ten-head vertical type insulator was developed to replace the older five-head horizontal type for non-quadded light gauge wire. In designing the new machine many improvements were incorporated. The old machines had been built to handle relatively strong paper and heavy wires, and studies indicated that to insulate finer wires successfully with lighter paper, also to run at high speeds without stretching the wire, and to apply a uniform wrapping without backlapping or folding over of the paper and with low breakage per pad the insulators must be rigid, the tension on the wires should be uniform and both supply and take-up mechanisms should operate smoothly.

The relative floor space per head for the ten-head machine including operator's space is about 60 per cent of that taken by the five-head machine but based on production the relative space per unit of production is about 50 per cent. The new machine runs at a head speed of about 3,000 R.P.M., carries a 12-in. pad of paper, and in general is a very substantial machine.

The insulating head, the vital part of the insulating machine, has undergone many changes to accommodate the thinner, narrower insulating papers. One of the most important of these has been to improve the tension mechanism which now consists of a very small multiple disc clutch actuated by a system of levers so that a very light but very uniform tension is applied at all times. This is not only making possible the use of smaller paper ribbon but may permit of changes in the composition of the paper with resultant cost savings. This head is shown in Fig. 8.

Another desirable feature in a paper insulating machine is a bare

wire detector as the insulation sometimes parts after passing through the sizing die or polisher as it is called, separates for a few inches and then picks up and goes on. Many electrical devices have been tried and practically all have the objection of high maintenance cost. A

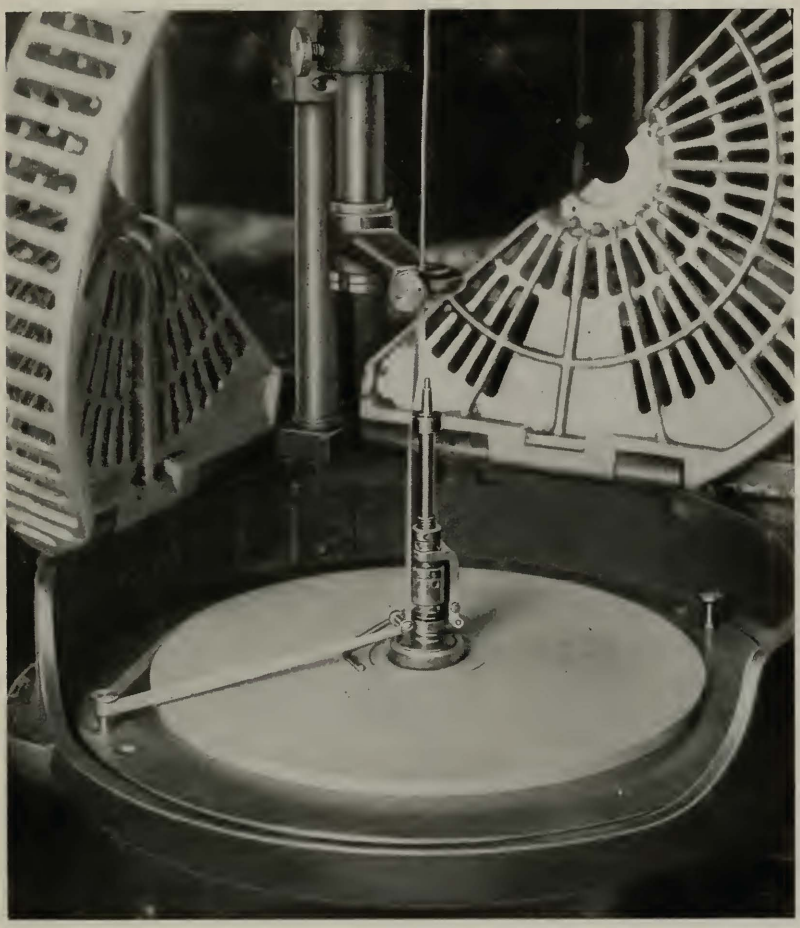


Fig. 8—Paper insulating head

very simple and effective remedy was the installation of a second polisher placed between the capstan and take-up spool which catches broken paper and pushes it back until the operator sees and repairs it.

The insulating machine used for heavy gauge wire is an eight-head machine built along the same general lines as the ten-head machine. This is illustrated in Fig. 9.

The method of splicing the copper wire is by means of a transformer, the low voltage side of which is equipped with clamps for holding the

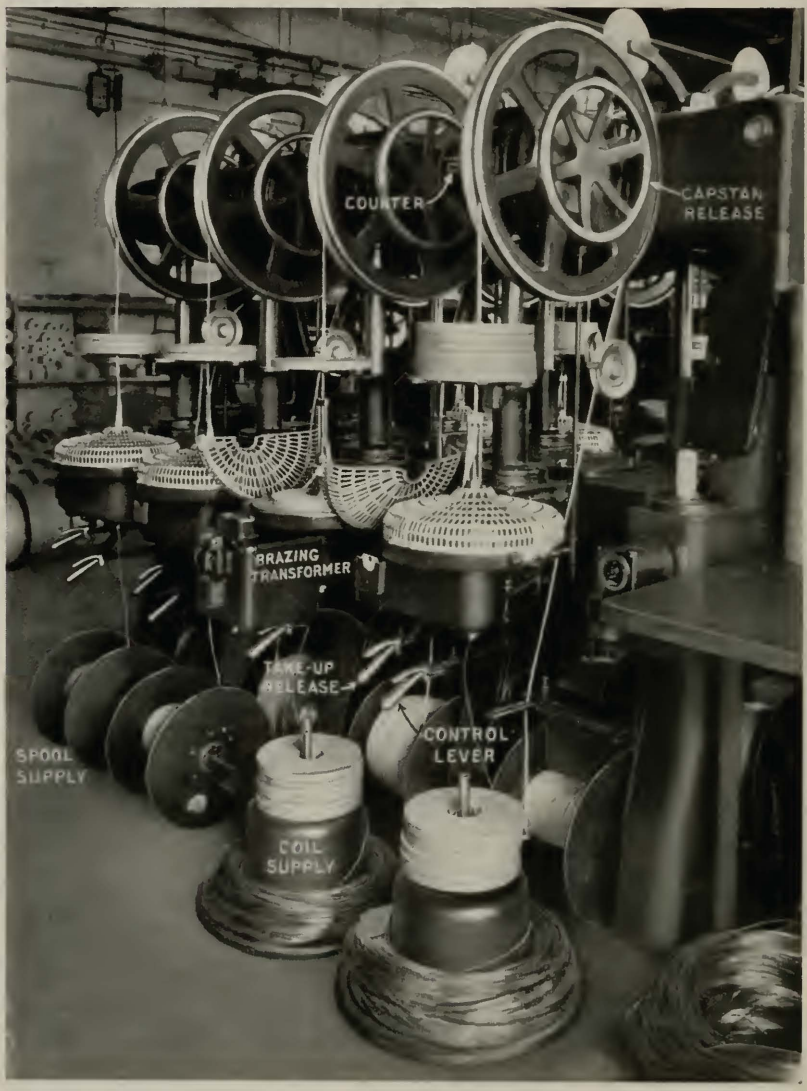


Fig. 9—Heavy wire insulator

two ends of wire which are butted together, heated by electric current and brazed by the application of borax flux and silver alloy solder. The transformer windings are so designed with low internal resistance

that, although different sizes of wire may be handled, the resistance of the wire between the clamps is so large in proportion to the total resistance that it automatically controls the current and prevents overheating of the wire.

Splices in the insulating paper are made by the application of a thin strip of gummed paper.

New twisting machines for non-quadded light gauge wire have been developed and these machines have some unique features which are worth a word of explanation. Fig. 10 shows schematically the old

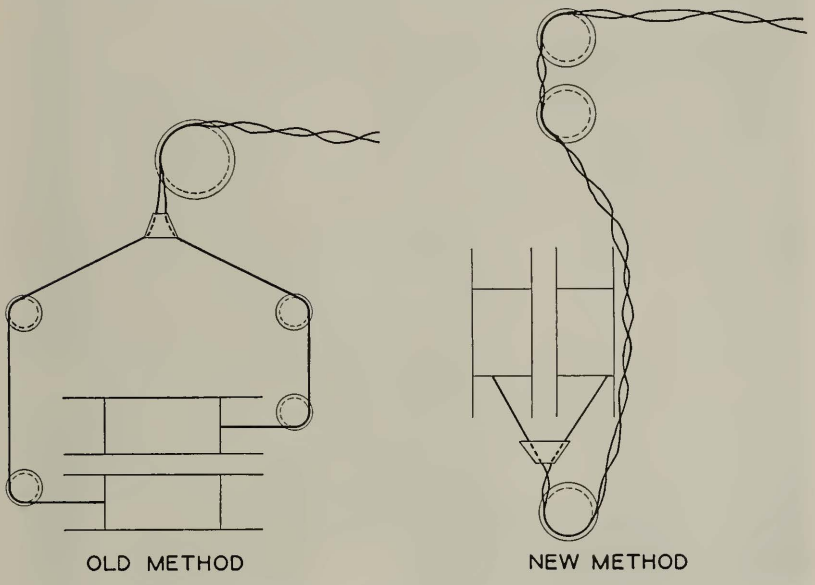


Fig. 10—Schematics of old and new type twisters

type of twister used ten years ago in which the two spools were placed with axes vertical inside of a flier which carried guide bushings through which the wire from the two spools was brought up to the center of the yoke and to the capstan. These machines operated at 500 R.P.M. and produced one twist per revolution. Assuming a 3-in. twist, the output would be about 125 feet per minute. In the new machines the spools are mounted side by side in a flier, the spools not revolving around each other, with axes horizontal, and the wire from each is taken off in a downward direction around a guide pulley and then up through the flier, around another guide pulley and to the capstan. With this arrangement two twists per revolution of the flier are produced and, as the machine is built to operate at 1,000 R.P.M., the out-

put for double the speed of the old machine is four times as great or about 500 feet per minute for a 3-in. twist. Additional features are

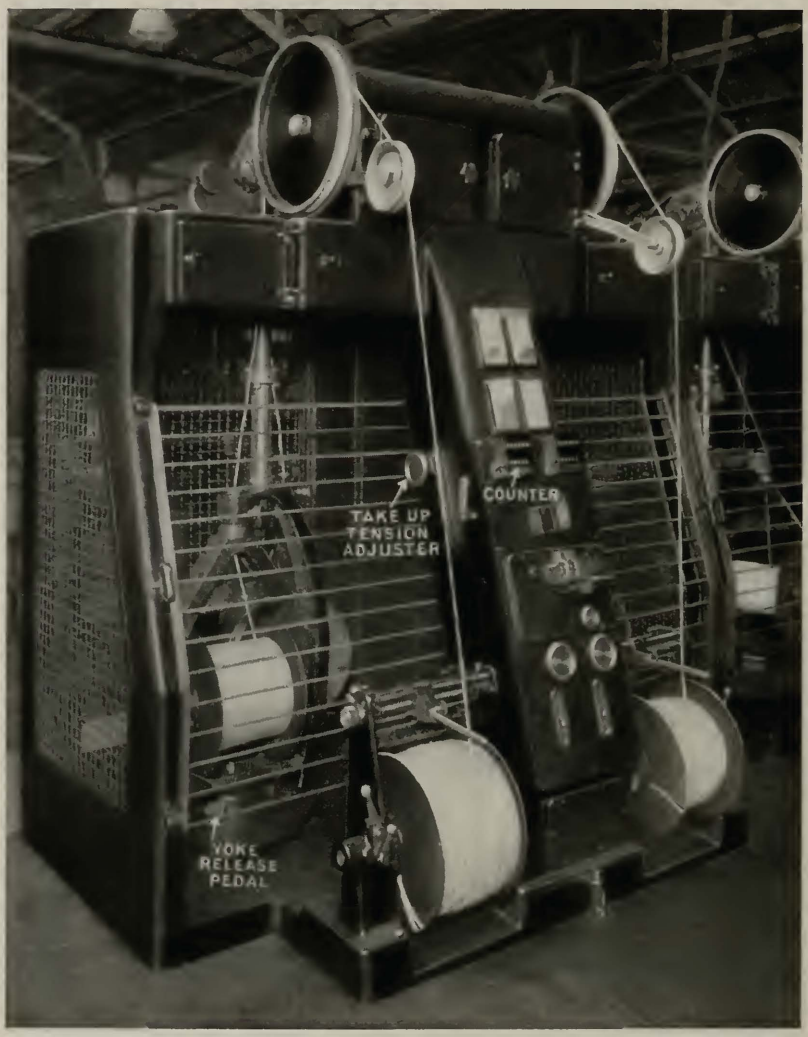


Fig. 11—Combined twisting and quadding machine

special tension devices to insure uniform tension on the wire and supports to assist in loading spools of wire into the yoke.

The twister for pairing and quadding heavy gauge wire in one operation is shown in Fig. 11.

Each spool, containing two conductors, is mounted in a yoke which revolves on its own axis to give the pair twist and the two yokes are revolved around each other to give the quad twist. This is accomplished by an arrangement of change gears from which can be obtained practically any length or direction of twist desired.

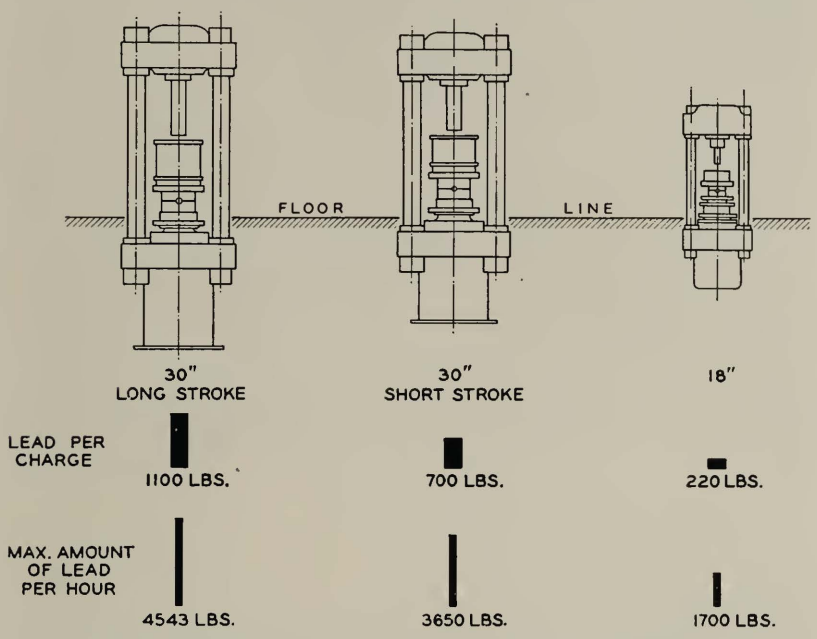


Fig. 12—Schematic showing relative increase in size of lead presses

Modern stranders follow the same general line as the older stranders but the whole design has been reviewed in detail with the view of strengthening and perfecting, and improved tension devices have been developed consisting of a tension arm actuated by the pair which in turn applies a brake to or removes it from the reel head. These are adjusted to give a tension of about three pounds per pair which causes no stretch and prevents over-running. With these, it is possible to run very fine wires at a minimum tension with a maximum smoothness of operation. The drums are gear driven and are capable of running up to 100 R.P.M.

After stranding, the cores are dried under vacuum to remove the moisture from the paper and then are covered with a lead alloy sheath.

It is necessary after the cable is removed from the vacuum drier to keep it in an atmosphere of a low moisture content until the lead sheath is applied. This was formerly accomplished by placing it in an oven

at a temperature of about 160 to 180° F. with a resultant relative humidity of not over 10 per cent. Cables maintained at this humidity would pick up very little moisture but in transit from the vacuum drier to the storage oven some moisture might be absorbed; also working in and out of these hot ovens was not particularly pleasant. Therefore, a method was developed for installing the vacuum driers in such a way that one end opens into an enclosed storage area in which the air is maintained at a temperature of about 100° F. and a relative humidity of less than 10 per cent until the cables are covered with lead. This temperature and humidity are obtained by cooling the incoming air to a dew point corresponding to the temperature and relative humidity desired and then passing it into the oven. A considerable engineering problem was involved in determining the heat given off by the vacuum driers and the hot cables and the additional moisture introduced by infiltration through walls, doors, etc.; also the relation between relative humidity, moisture content of paper and electrical characteristics presented a most interesting field for study.

The method outlined above has proved very satisfactory as the cables do not absorb enough moisture to affect their electrical properties and the conditions in the storage area are not unpleasant; in fact, during the summer time they are somewhat more agreeable than the outside air during periods of high humidity.

The process of applying lead sheath to cable is one which has not undergone any change in principle since sheath was first applied directly to the cable instead of cable being pulled into it. There have been, however, a number of developments tending to improve the quality or increase the output.

In covering a large cable something more than half of the total time of one cycle of operation is taken up by filling the cylinder with lead and cooling under pressure to the point where it can be extruded. The tendency, therefore, has been to build presses with larger lead containers in order to increase the time of extrusion relative to the total cycle.

The diagram (Fig. 12) shows schematically an early type of press, one which was considered standard a few years ago, and one of the presses designed and built recently. Underneath each press is a figure showing the lead content per charge and the relative amount of lead extruded per hour by each of the three presses.

As will have been noted from the diagram, the stroke of the newest type of presses is about one foot longer than that of the former presses although the diameter of the lead container and the diameter of the water ram are the same.

The pressure for operating these presses is furnished by a hydraulic pump, pumping water at six thousand pounds pressure per square



Fig. 13—Electric tractor and trailer for handling cable reels.

inch. Presses were formerly connected to four plunger vertical type pumps, but it was found that more water could be used with the large



Fig. 14—Insulating machines

sizes of cable and, therefore, new pumps were built with six plungers, giving a proportionally greater output. The diameter of the lead ram

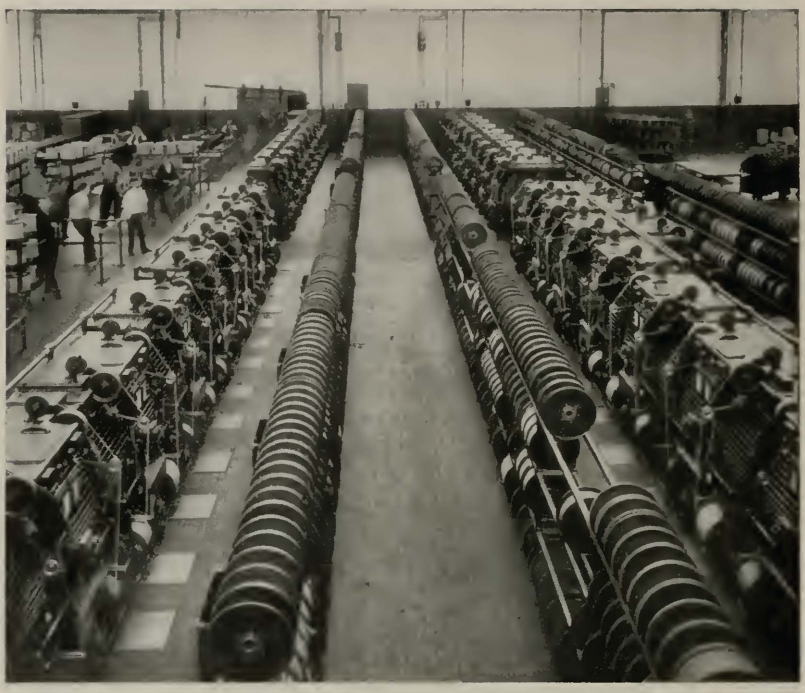


Fig. 15—Twisting machines

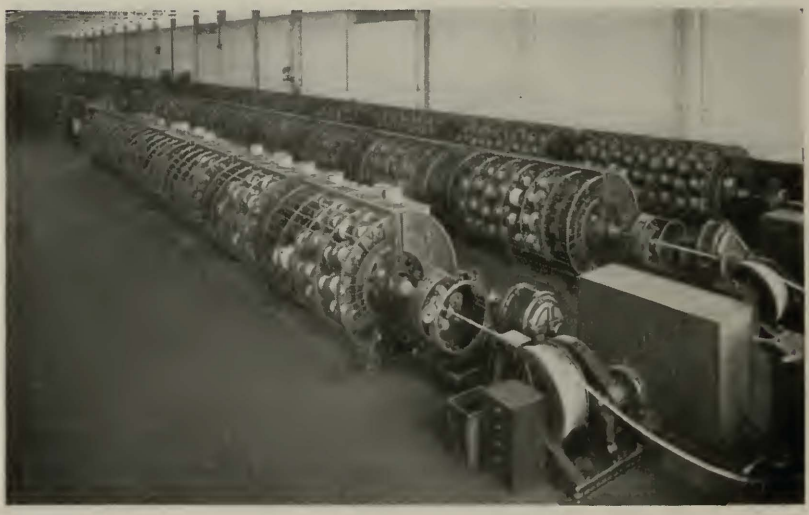


Fig. 16—Stranding machines



Fig. 17—Lead press equipment



Fig. 18—General view of reel yards

is one third that of the water ram, so that the pressure on the lead during extrusion is about 54,000 pounds per square inch.

Aside from increasing output many studies have been made to determine the exact mechanism of lead extrusion, the relative flow of lead in different parts of the extrusion block, the effect of application of heat at different points, etc.

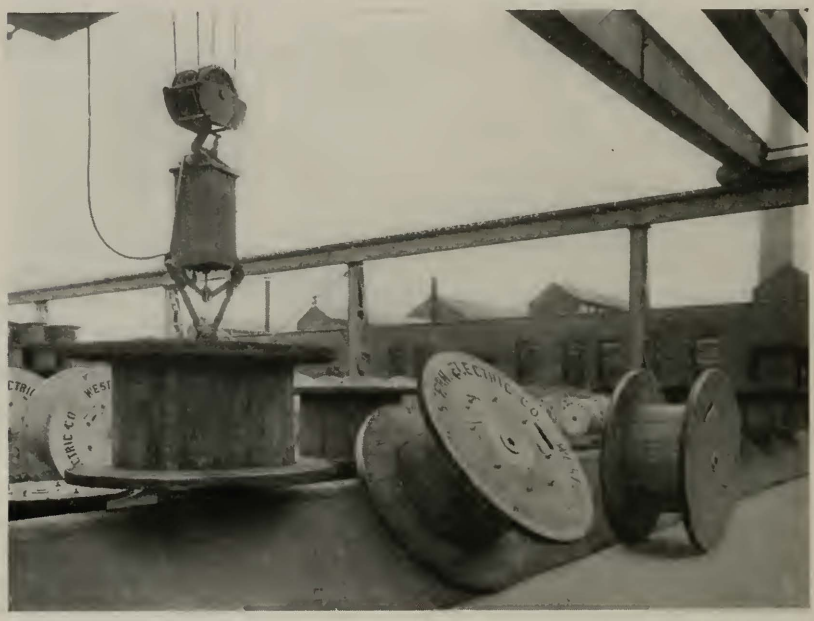


Fig. 19—Delivery of empty reels from yard

An interesting experiment consisted in filling an extrusion block with layers of different colored waxes and noting their flow under pressure. This gave valuable data as to the proper contour of the extrusion chamber.

The concentricity of sheath is affected not only by the contour of the extrusion chamber but also by the manner in which heat is applied; and thickness is affected by temperature and speed of extrusion so that the human element is an important factor, and it is necessary to have thoroughly trained and reliable operators on this kind of work. Temperature indicators are used to show die block temperatures and the temperature of the molten lead is automatically controlled and recorded.

Testing, Storage and Shipment

Handling of lead-covered cable on reels, the total weight of which runs from one to five tons, is a very distinct problem. This handling from press to test is done by a crane which picks up the reels and carries them to the place where they are to be tested for insulation resistance, capacitance, dielectric strength, etc.

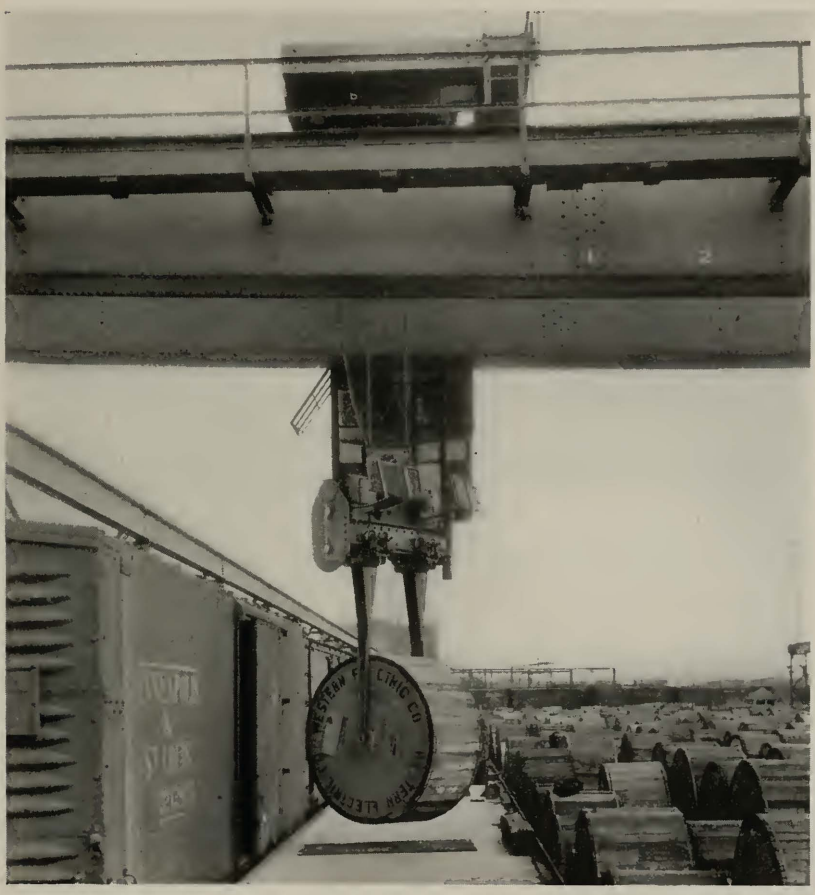


Fig. 20—Crane placing reels on loading platform

After the cables are tested, the ends are sealed and wooden lags are fastened around the periphery of the reels after which the cables are taken to a storage yard until the customer's order is completed, at which time they are shipped. A special tractor and trailer, Fig. 13, has been developed and substituted for manual handling.

Handling cables from the reel yard to the loading platform was a very serious problem, particularly in the winter during snow storms. This was taken care of by the installation of overhead cranes for picking up reels and placing them on the platform.

The lifting mechanism for empty reels consists of a solenoid-operated plunger controlled by the crane operator. The reels are turned on the side, the plunger inserted in the bushing and the operation of the solenoid throws out two lugs which prevent the plunger from being withdrawn and lift the reel. When the reel is to be released, it is put down on an inclined surface which turns it back on to its flanges. This method of lifting empty reels permits them to be stacked one on top of the other and saves storage space.

The lifting mechanism for full reels consists of two side arms with lugs moved horizontally by means of a double-threaded screw and a motor controlled by the crane operator. With this device the crane operator can pick up and put down any reel without the assistance of a ground man.

Figs. 14 to 20 show insulating, twisting and stranding machinery, lead presses and cable reel yard with cranes and special lifting equipment for both empty and full reels.

The methods of cable manufacture are ever changing. What has been described as strictly up to date today will, doubtless, on account of new developments be superseded by new methods, new equipment and new designs, so that the Cable Plant of the future will be different from and more efficient than that of the present.

Bridge for Measuring Small Time Intervals

By J. HERMAN

SYNOPSIS: A bridge circuit for measuring time intervals from about one ten-thousandths of a second up to several seconds is described and its operation explained. The device is fairly accurate and easy to operate and gives the results of measurements in fractions of a second directly. Its calibration can readily be determined mathematically since it is dependent only upon the values of certain capacities and resistances used in the measuring circuit.

TO the large family of measuring devices making use of certain principles of electrical balance there has recently been added a new member. This measures the elapsed time between the opening or closing of one set of contacts, and the subsequent opening or closing of another set of contacts, the agency employed for operating the contacts being immaterial. The particular form of the device described below, was designed primarily for use in adjusting the operating and releasing times of the voice-operated switching relays at the terminals of the transatlantic radio telephone circuit. In this form or with minor changes, it is applicable to the measurement of intervals of time in the operation of a large variety of other types of apparatus.

The new time measuring device is simple and easy to operate, its operation consisting merely of opening and closing a key repeatedly and securing a balance by observing a meter. The balance is secured by turning one or more dials which are calibrated in fractions of a second.

A range of measurements extending from about one ten-thousandths of a second up to several seconds is readily obtainable and an accuracy of measurement to within $\pm$ 1 per cent can probably be realized over the greater part of this range if sufficient care is taken in the design of the circuit and the selection of the apparatus. In a fairly rugged type of bridge, now in commercial service (See Figs. 3 and 4), which covers the range from one ten-thousandth of a second to one second, and in which little attempt was made to secure a high degree of sensitivity, the results of measurements are accurate to within $\pm$ 5 per cent for time intervals down to about five thousandths of a second. For time intervals below this value the accuracy decreases rapidly, due partly to the fact that the smallest step provided for on the dials is one ten-thousandth of a second and partly due to the effect of variations in the operating time of the relays used in the bridge.

Because of its simplicity and accuracy, the bridge is especially valuable for making a series of time measurements to determine the

best adjustment and the most desirable circuit condition for the operation of a relay or similar device. With the bridge, this requires very little time, especially since the results of the individual measurements are immediately available to guide the work. With the oscillograph, several hours may be required and the results obtained are available only after developing and analyzing the oscillograms.

The calibration of the bridge is determined by the values of certain capacities and resistances in the measuring circuit. These may usually be selected with sufficient accuracy during manufacture so that the bridge requires no further calibration after it has been constructed. In fact, it has been found practicable to design the bridge so that the steps on a standard decade resistance box correspond to decimal fractions of a second.

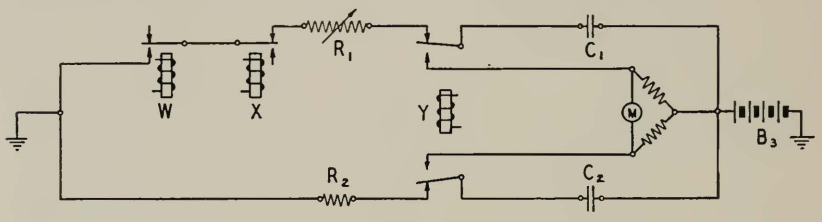


FIG. 1

The principles underlying the measurement of an interval of time will be explained in connection with Fig. 1. Two condensers C_1 and C_2 of unequal capacity are charged from a common battery B_3. The condenser C_1, which has a larger capacity than C_2, is charged through an adjustable high resistance R_1 during the time elapsing between the operation of relay W and the subsequent operation of relay X. This elapsed time is the interval of time to be measured and the charge accumulated on the condenser is an accurate means for doing so. The second condenser C_2 is used merely for comparison purposes. It is charged through a fairly low resistance R_2 and acquires its full charge in a relatively small interval of time.

After the completion of the charging interval, relay Y is operated and the two condensers are discharged simultaneously through a differential meter circuit. If the two charges are equal, the meter will show no deflection, but if they are unequal, it will show a momentary deflection, the direction of which will indicate whether the charge on C_1 is too high or too low. By repeating the charging and discharging process a few times and adjusting the value of resistance R_1 in series with the first condenser, the charges on the two condensers can be made equal. When this condition is obtained, the interval of time during

which the charging took place may be determined from the value of the high resistance.

The relationship between the interval of time of charging and the value of resistance required to make the charges on the two condensers equal is a direct proportion. This will be obvious from an inspection of the general equation for the charge at any instant on a condenser which is being charged through a high resistance. The equation is

$$q = Q(1 - e^{-(t/CR)}), \tag{1}$$

where

q = charge at time, t,

t = elapsed time in seconds since the charging began,

Q = final or maximum charge on the condenser,

C = capacity of condenser in farads,

R = resistance in ohms in series with the condenser,

e = base of Naperian logarithms.

Since q is always made equal to the charge on the comparison condenser and since the charging battery is common to the two condensers, therefore, using the symbols shown in Fig. 1, C_2E may be substituted for q and C_1E for Q. The equation then becomes

$$C_2 = C_1(1 - e^{-(t/C_1R_1)}). \tag{2}$$

As mentioned above, the two condenser capacities are kept constant. Therefore, any change in t requires a proportional change in R_1 in order to satisfy the equation.

Fig. 2 is a schematic circuit diagram of the complete time measuring bridge showing the manner in which it may be connected to a representative type of circuit to be tested (shown by dotted lines). The symbols used to designate the various circuit elements in this figure are the same as those used in Fig. 1 and since the principles of operation have already been explained in connection with the latter figure, a comparison of the two figures will aid materially in understanding the detailed circuit arrangements of the bridge.

As shown in Fig. 2 the bridge is arranged to measure the operating time of a voice operated switching device consisting of a detector and a relay Z in the output circuit of the detector. The input circuit of the detector is connected to the output circuit of an oscillator and may be opened or closed by contacts on one of the bridge relays (relay W). This relay is under the control of key K which, when closed, causes relay W to operate and complete the oscillator connections to the

detector thereby initiating the operation of relay Z. At the same time, another set of contacts of relay W close the charging circuit of condenser C_1 thereby permitting the charging of this condenser through the high resistance R_1. These conditions remain unchanged until the armature of relay Z has reached its m contact and caused the operation

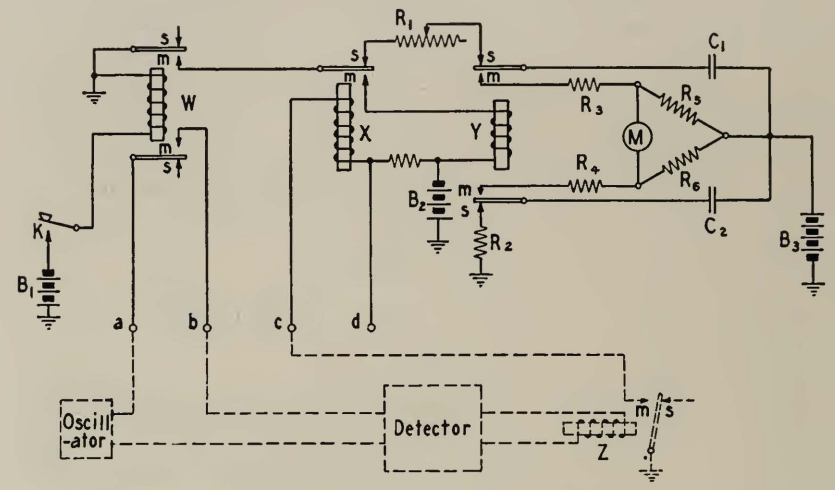


FIG. 2

of relay X. The latter relay first opens the charging circuit of condenser C_1 and then causes the operation of relay Y. The operation of relay Y discharges the two condensers C_1 and C_2 through the differential meter circuit composed of the meter M and the two equal resistances R_5 and R_6. Additional resistances R_3 and R_4 are connected into the discharge circuit to limit the discharge current and prevent sparking at the relay contacts.

If the meter shows no deflection at the instant of discharge, it indicates that the two charges are equal and that the operating time of relay Z is as indicated by the value of resistance R_1. If the meter shows a deflection, the key K should be opened and closed repeatedly and the resistance R_1 adjusted until the meter shows no deflection.

In order to prevent a quick double deflection of the meter when a balance has been reached, it is necessary that the time constant of the two branches of the discharge circuit be approximately alike. For this reason, the ratio $R_4 + R_6$ to $R_3 + R_5$ of the discharge circuit resistances should be approximately the same as the ratio C_1 to C_2 of the condenser capacities.

The releasing time of relay Z; that is, the time required for the armature to leave its m contact after the oscillator is removed from the detector, may be measured by making a few simple changes in the

connections. The oscillator is connected directly to the input of the detector and the terminals *a* and *b* of the bridge are connected across the oscillator output. Contact *m* of relay *Z* is connected to terminal *d*, and terminal *c* is grounded. The closing of key *K* will then cause a short circuit to be placed across the output of the oscillator, thereby initiating the releasing of relay *Z*. As soon as the latter occurs, a short circuit is removed from the winding of relay *X*, allowing this relay to



FIG. 3

operate and open the charging circuit of the condenser C_1. Except for the differences noted, the operation of the bridge is the same as described above.

Other conditions of measurement, for example, a measurement of the time required for the armature of relay *Z* to reach its *s* contact after a short circuit is applied to the output of the oscillator will be obvious from the two examples given. It should also be obvious that a battery and suitable resistances may be substituted for the oscillator and detector and measurements made in this direct-current circuit in the same manner as above.

Assuming that the bridge has been properly calibrated from theoretical considerations of the constants of the measuring circuit, there are two possible sources of error in the results obtained. These errors are due to the time of operation of the two switching relays W and X. In the case of relay W an error will be caused if its two sets of contacts do not close simultaneously. In the case of relay X the source of



FIG. 4

error is due to the time required for the relay to open the charging circuit at its s contact at the end of the operating or releasing time which is being measured. The combined error due to the operation of the two relays may actually be determined by means of the bridge itself and substracted from the results of the measurements. To determine the error, terminal a is connected to ground and terminal b to terminal c. The key K is then operated in the normal manner and a balance secured on the meter. The readings of the dials will then indicate the error of the bridge.

In order to avoid the necessity of correcting the measurements for the error in the bridge, a value of resistance corresponding to the error may be connected permanently in series with R_1. This should be made variable if a high degree of accuracy is desired, because the operating time of the two relays in the bridge may change slightly from day to day and small readjustments of the resistance will, therefore, be necessary.

The sensitivity of the bridge, that is, the ease with which small intervals of time can be distinguished by the deflection of the meter, is dependent upon the sensitivity of the meter, the voltage of the common charging battery, the value of capacity, and the ratio of the capacity of the large condenser to that of the small condenser. With a particular type of meter, the larger the voltage, the condenser capacities, and ratio of condenser capacities, the more sensitive will be the bridge.

Although the chief use for the bridge up to the present time has been in connection with voice operated relay devices, its future use will undoubtedly be extended to other fields. Some indication of the uses to which it might be put may be obtained from the following suggested applications.

1. For measuring the functioning times of electromagnetic switching arrangements in various kinds of communication and signaling systems.
2. For measuring propagation time at different frequencies over telephone circuits and "lag" in telegraph instruments and circuits.
3. For studying the operation of a machine with a view to improving it by measuring the speed of operation of various parts and the relative time of operation of certain parts with respect to other parts. Electrical contacts would have to be provided temporarily at suitable places on the machine.
4. For maintaining the proper adjustment of time-limit over-load relays, circuit breakers, etc., on power circuits.
5. To determine the rate of acceleration of motors or other machinery by employing a suitable centrifugal contact arrangement.
6. In psychological tests for determining whether the time of response of a person to a particular signal is above or below a required value.

Numerous other applications to laboratory and field work might be suggested where such a time measuring device could be employed to considerable advantage. The cases mentioned above are not necessarily the most practical but they serve to illustrate the capabilities of the device as described or when provided with simple modifications.

A Method of Rating Manufactured Product

By H. F. DODGE

SYNOPSIS: This paper outlines a method of rating manufactured product. In the particular form here described, the rate has been found very useful for measuring the quality of communication equipment and materials entering the plant of the Bell System. While the primary object is control of quality of finished product, it is proving useful for measuring the workmanship of individual operators and groups of operators engaged in similar production work. Particular attention is directed to the statistical aspects of the rate to show how it can assist in controlling quality.

GIVEN a product whose quality is dependent on a number of diverse characteristics, the following questions and others similar to them frequently require answers. Has quality been satisfactorily controlled? Is there any general trend in quality either upward or downward? How does current quality compare with that of a year ago?

These are questions of importance to the manufacturer. Qualitative answers can often be given on the basis of general knowledge by those familiar with the details of manufacturing performance but such answers tend to be inaccurate or biased. What is often wanted is some statistical index based on quantitative data, a figure which balances the favorable features against the unfavorable to give an overall picture of quality *on the average.*

To get such a picture does not in general require special data. The detailed data obtained in the course of routine inspection, while often used only for the immediate purpose of determining the satisfactoriness of individual lots of product, are just what is needed for the present purposes. These inspections are critical examinations of the features that are essential to proper operation of the product in service. Hence the results are a measure of quality. There are of course many possible ways of classifying and combining this quantitative information, some of which are more efficient than others. The problem is to set up a method of handling the data in a way which will paint as clear a picture as possible of the overall quality.

The rate here described has been found very useful for measuring the quality of communication equipment and materials entering the plant of the Bell System.[1] It recognizes and takes account of the relative seriousness of different types of defects found in the course of

[1] This method of rating is being used extensively by the Manufacturing Department of the Western Electric Company where some of the features outlined in this paper originated.

inspection. For convenience, the rate is made a relative figure which incorporates the features of index numbers used by the economist. Just as the index numbers of cost of living, wages, corn production, etc., indicate current conditions relative to some reference condition as

$$\text{Index Number} = 100\,\frac{\text{Current Cost of Living}}{\text{1914 Cost of Living}},$$

so does the rate reflect current quality relative to that of a selected standard of reference. One of the features of the rate is its assistance in controlling quality, its provision of means for discriminating between chance and non-chance variations from the quality level which should currently be expected.

Character of Inspection

As in other fields much of the inspection work on telephone products consists of critical examinations of essential features to determine whether or not the units of product conform with specification requirements. This is done:

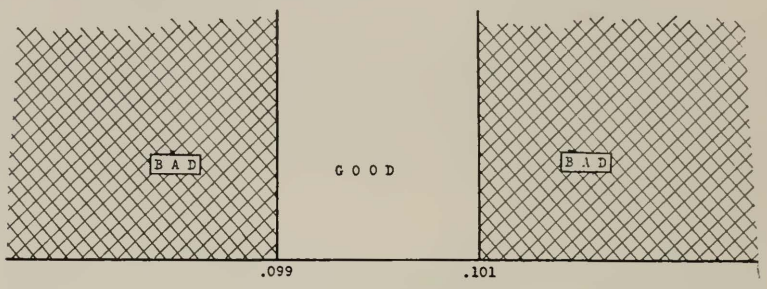
(1) By visual examinations in which obvious defects of material or workmanship are discovered by eye.
(2) By using "Go" and "No Go" gauges or their equivalent, which determine whether a unit does or does not conform with a requirement, or
(3) By using measuring instruments which reveal the numerical magnitude of the characteristic for each unit tested.

To illustrate the last two kinds of inspection, the specification requirement for the capacity of a type of condenser is "not less than .099 microfarad and not more than .101 microfarad." Inspection may be done by the "Method of Attributes," using a test set which shows merely whether the capacity of a condenser is inside or outside of the limits, or by the "Method of Variables," using an indicating or recording meter to show the numerical value of capacity for each test. In these two cases the data, if tabulated, would appear as in Fig. 1. Inspection data used for rating come in both varieties.

Items which Enter the Rate

Commercial measurement of quality by inspection usually consists in a comparison with stated requirements. Starting with the design and a knowledge of what can be accomplished in the shop, allowances for variations in materials, dimensions and salient properties are established in specifications. The aggregate of specification requirements constitutes a standard of quality which the manufacturer holds

before him as an upper limit of attainment. To him, perfect performance is 100 per cent conformance with requirements and the resulting product he regards as of "perfect quality." The rate encompasses this narrow viewpoint of quality and measures the success to the manufacturer in living up to this adopted standard.



Capacity (in Microfarads)

Inspection by Method of Attributes		Inspection by Method of Variables	
Condenser No.	Observation	Condenser No.	Observation
1	Good	1	.0991
2	Good	2	.1006
3	Bad	3	.0985
4	Good	4	.0995
—		—	
—		—	
—		—	
121	Good	121	.0999
122	Bad	122	.1013
123	Good	123	.0994

Fig. 1—Two methods of measuring the quality of condensers in respect to capacity

The only items which enter the rate are the "defects," i.e. failures to meet requirements, found in the course of inspection. Experience has shown that percentage non-defective, the ratio of perfect parts to the total parts, while useful for certain classes of investigation, is not a very satisfactory yardstick for measuring quality of complex products. This factor fails to take into account two important things:

(1) Defects of different kinds are not equally serious.
(2) Defects of the same kind vary in seriousness according to the degree of departure from specified limits.

Thus a failure to meet a major requirement should have greater weight than a failure to meet a minor one and in like manner the degree of imperfection of a given kind should be taken into consideration.

The rating method recognizes such gradations in seriousness by making use of a system of weighting defects.

Method of Weighting Defects

The seriousness of a defect is judged from the standpoint of the consumer. A defect, if allowed to get into service, means trouble in one form or another, and trouble costs money. Seriousness depends fundamentally upon the evaluation of the loss or expense that would be incurred by using the defective unit. The determination of exact costs of trouble is generally not possible but these costs or, better, the relative costs can be estimated. Such estimates may be based on past experience, judgment, engineering knowledge of service requirements, complaints received from consumers and available information on costs associated with past troubles in service.

A standard set of classes is adopted for defects associated with a given kind of product, the classes being ordered in seriousness and each sufficiently well defined to make the business of classification a fairly simple and uniform process. The following four-fold classification has been found satisfactory for many kinds of telephone products.

Class "A" Defects—Very serious.

Will render unit totally unfit for service.

Will surely cause operating failure of the unit in service which cannot be readily corrected on the job, e.g. open induction coil, transmitter without carbon, etc.

Liable to cause personal injury or property damage.

Class "B" Defects—Serious.

Will probably, but not surely, cause Class "A" operating failure of the unit in service.

Will surely cause trouble of a nature less serious than Class "A" operating failure, e.g. adjustment failure, operation below standard, etc.

Will surely cause increased maintenance or decreased life.

Class "C" Defects—Moderately serious.

Will possibly cause operating failure of the unit in service.

Likely to cause trouble of a nature less serious than operating failure.

Likely to cause increased maintenance or decreased life.

Major defects of appearance, finish or workmanship.

Class "D" Defects—Not serious.

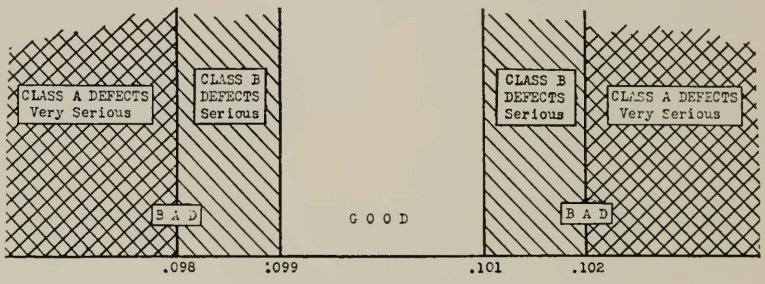
Will not cause operating failure of the unit in service.

Minor defects of appearance, finish or workmanship.

It should be pointed out that the number of classes to be used is arbitrary. Two classes, major and minor, may be sufficient for some

relatively simple products. The number of classes that can logically be used in any case depends upon the accuracy which can be attained in making estimates of relative seriousness.

Before proceeding further it may be well to indicate how the defects for features which are inspected as "variables" are weighted. Take the illustration accompanying Fig. 1. Any failure to meet the commercial limits of .099 and .101 microfarad will result in irregularities in transmission such as the distortion of the words spoken over a telephone line. The greater the departure from these limits the greater is the seriousness from a service standpoint. Strictly the weight for a defect should depend upon the degree of its departure from a limit but the desired result can be approximated to a satisfactory degree of accuracy by classifying the defects into two or more classes. To illustrate, assume two classes as indicated in Fig. 2. Defects falling within the



CAPACITY (in microfarads)

Fig. 2—Classification of defects for variable characteristics

ranges .098 to .099 and .101 to .102 are serious and can be considered as Class "B" defects in a four-fold classification while defects outside of the two outer limits .098 and .102 are Class "A" defects and can be weighted as such.

COMPUTATION OF THE RATE

A defect is weighted by assigning to it a number of "demerits." For a given kind of product each class of defects has a specified weight. Since the *relative* weights are alone of importance, the scale of demerits may be chosen arbitrarily.

The unit of measurement in the rating plan is "demerits per unit." [2] This factor is the simple sum of the demerits per unit contributed by the different types of defects found in inspection.

[2] The "unit" is commonly a physical unit of product such as a piece part, a partial assembly or a finished unit of apparatus or equipment. Exceptions to this rule have been found desirable for certain complicated types of product, such as switchboard sections ori nstalled central office equipment, in which cases the unit may be a natural element of a physical unit such as a soldered connection, a circuit, etc.

$$\text{Demerits per unit} = \frac{w_1 d_1}{n_1} + \frac{w_2 d_2}{n_2} + \cdots \qquad (1)$$

for all types of defects, where

w_1, w_2, etc. = weight (demerits per defect) for defects of type 1, 2, etc.

d_1, d_2, etc. = number of type 1, 2, etc., defects, and

n_1, n_2, etc. = number of units inspected for type 1, 2, etc., defects.

Instead of using equation (1) directly for indicating quality, it has seemed desirable to establish a rate which by its numerical magnitude gives an immediate indication of whether the quality is better or worse

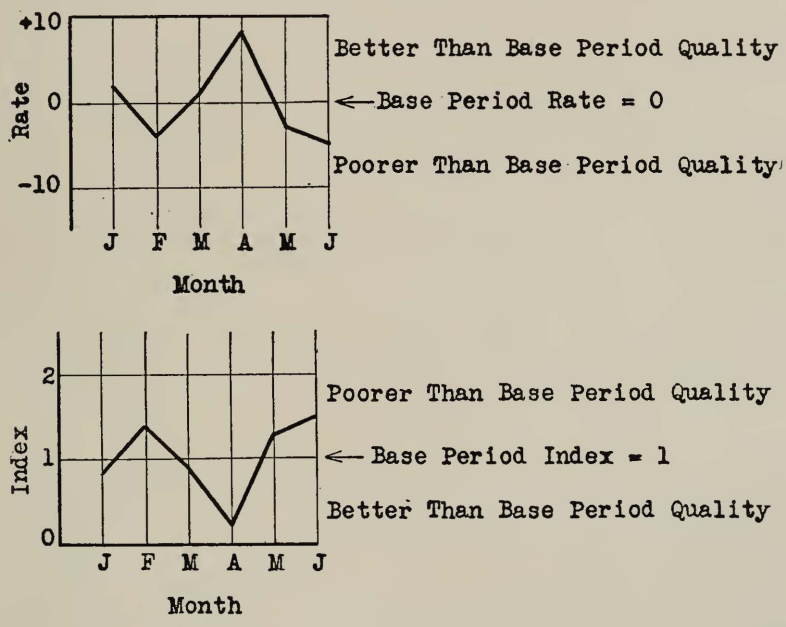


Fig. 3—Relation between index and rate for a given product and period of time

than that of some easily recognized reference condition. Resorting to methods commonly used in constructing index numbers, we therefore select a base period during which the manufacturing conditions and inspection methods were known to be essentially the same as at the present time, determine the demerits per unit for the representative data of that period, and set up the following index:

$$\text{Index} = \frac{\text{Current Demerits per Unit}}{\text{Base Period Demerits per Unit}}. \qquad (2)$$

Since the demerit is an element of badness, this index increases in magnitude as the quality grows worse as shown in the upper chart on Fig. 3. It is preferable to have the rate high when the quality is good and low when the quality is bad. This has been taken care of by using the factor (1 − Index) in the rate equation,

$$\text{Rate} = 10\,(1 - \text{Index}), \tag{3}$$

where the factor 10 is introduced merely to make a convenient scale. This gives a rate of +10 for a product of perfect quality (i.e. no defects found in the material inspected), a rate of zero when current quality is the same as the average for the base period, and a negative rate when current quality is poorer than that of the base period. This equation, as portrayed by the lower chart of Fig. 3, is merely a numerical way of saying "better than" or "poorer than" base period quality and it also tells how much better or poorer.

The choice of base period rests on judgment and knowledge of conditions and must be made with the eyes open. To take care of evolutionary changes in manufacturing conditions for telephone products it has been found desirable to use a *moving* base period [3] of not longer than five years. The use of a somewhat extended period where possible has the advantage of stability in that it tends to smooth out the high and low spots resulting from temporarily abnormal conditions of production such as are liable to recur in the future. The magnitude of the *base period demerits per unit* thus establishes a reference level for quality under *average* conditions.[4]

Quality-Control Feature of the Rate

Rates obtained from week to week or from month to month are used to indicate whether quality has been controlled. If manufacturing conditions are steady and everything is running smoothly, some definite value of rate can be *expected*. But even with a perfectly controlled process, there will be fluctuations above and below the expected rate value, fluctuations resulting from the effects of a large number of causes over which the manufacturer has no control.

[3] By a moving base period of 3 years is meant the three years just preceding the current year. With a moving base period the standard of reference (the denominator of the index) will change slightly at the beginning of each year as one year is dropped and a new one, the preceding, is added to the base.

[4] The average of past experience is sometimes a suitable estimate of *expected* quality but its indiscriminate use for this purpose is to be avoided. For products which are reasonably well controlled this estimate will often serve satisfactorily. Primarily the denominator of the index is a magnitude chosen to represent some standard of reference. The numerical rate obtained at any time reflects quality relative to the standard. It is not essential to the rate that the expectancy feature be stressed in this connection. Expectancy is, however, of importance to the control limits discussed in the subsequent paragraphs.

How low does a rate have to fall before lack of control is indicated? Does a rate of -10 signify that something abnormal has happened? The following discussion gives a method which can be used to detect lack of control.

First of all we must determine the value of rate to be expected, i.e. establish a norm for expected quality. Past experience can usually be used as a guide for this purpose. If the average quality during the base period is considered satisfactory as an estimate of expected quality under current conditions, then the expected rate is 0. If only a portion of past data is judged suitable for this purpose, then the rate figure corresponding to the selected data is the expected value.

The method of establishing limits of expected variation for the rate makes use of statistical methods which have been described elsewhere,[5] but will be briefly reviewed. If the current rate deviates from the expected rate by an amount which is greater than can be attributed to chance, this will be taken as an indication of lack of control.

Just how chance enters the discussion will perhaps be better understood from the following. Each unit of product is the physical result of fashioning and combining various materials by a large number of manual and mechanical operations and processes. Every element in the production process which contributes to the final detailed character of a unit can be considered as a cause. Now the ideal state of affairs, purely conceptual to be sure but nevertheless one which is the goal in all attempts to secure greater uniformity of quality, is one in which each of the elemental causes or groups of causes (affecting a particular trait of the product) functions continuously in the same manner to produce a given elemental effect in the direction of defective quality. Considering overall quality, one group of manufacturing causes is responsible for one type of defect, another group for a second type, etc. The aggregate of these many causes which cooperate to mould the product may be considered as a system of causes. When the concept of constancy-with-time is associated with all of the causes, the system is spoken of as a "constant system of causes," [6] i.e. one whose tendency toward defective quality does not change with time. Product turned out by such a system will be referred to as "uniform product."

For product which is uniform in this sense the rates obtained week

[5] "Quality Control Charts," by W. A. Shewhart, *Bell Sys. Tech. Jour.*, Vol. V, pp. 593–603, October, 1926.

[6] "Application of Statistics as an Aid in Maintaining Quality of a Manufactured Product," by W. A. Shewhart, *Jour. Am. Stat. Ass'n*, Vol. XX, pp. 546–548, December, 1925. It should be noted that the system of causes associated with the *data* used in rating is all-inclusive, encompassing the causes which are responsible for inaccuracies of measurement introduced by inspection as well as the manufacturing causes which affect actual quality.

by week or month by month will fluctuate around some average value according to the laws of chance. For example, in the manufacture of selectors assume conditions are such as to give uniform quality with an expected rate of 0. The rates for batches of selectors turned out weekly will fluctuate about Rate = 0, the range of variation depending on the number produced each week. A week's output can be regarded merely as a sample of the product which this system of causes would turn out if it were allowed to function in the same manner for an

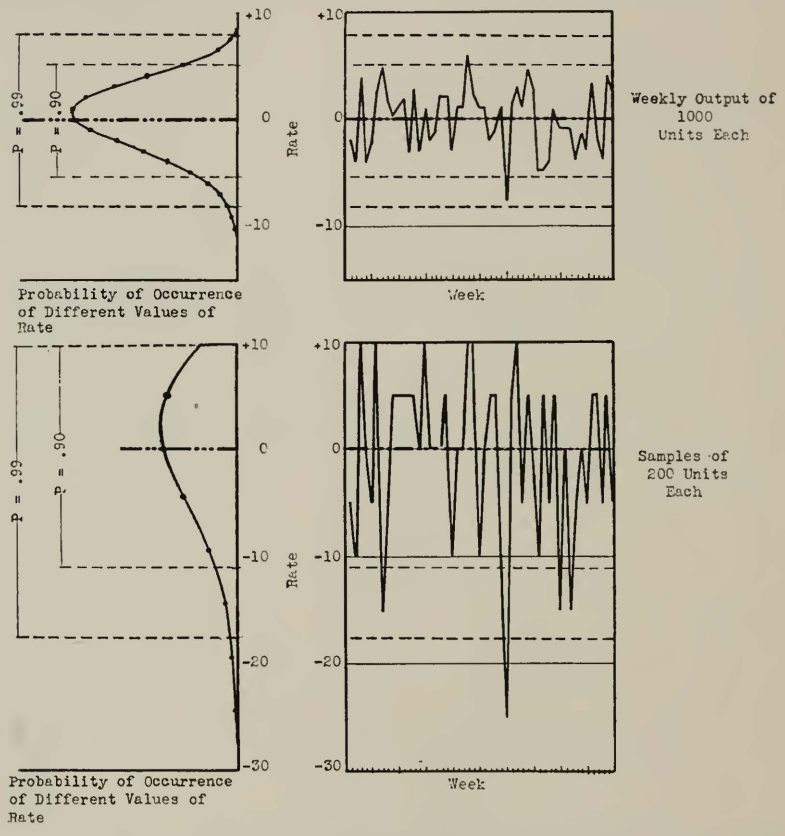


Fig. 4—Typical fluctuations of rates for uniform product whose expected rate = 0

indefinite length of time. The distribution of weekly rates is the same as would be obtained in an ordinary sampling experiment by drawing samples from an infinite warehouse of thoroughly mixed selectors having an average quality represented by Rate = 0. If inspection consists in examining only a percentage of the selectors manufactured, this will be merely equivalent to taking smaller quantities of selectors

from the warehouse and the resulting rates will be spread out more widely around 0 than when the entire product is inspected. These results are exemplified by the two diagrams in Fig. 4.

The probability curves of Fig. 4 represent the basis for setting control limits. The area under the curve between any two limits divided by the total area represents the probability that a single rate will fall between these limits. For a probability of .99 we can say that if the product is controlled at a level corresponding to the expected rate (zero in the illustration) then the chances are 99 in 100 that the current rate will fall within the limits thus established and only 1 in 100 that it will fall outside the limits.

For any product the spread between the limits is governed by two factors, the number of pieces inspected and the value of the above probability. It is necessary therefore to make an arbitrary choice of probability, a choice which will depend on the use to be made of the rate.

The control lines are used primarily to distinguish between those variations which may be attributed to chance causes and those which are more probably the result of some significant change in manufacturing conditions, either production or inspection, and therefore worthy of investigation. The criterion of the suitability of the limits chosen is the percentage of cases falling outside of the limits which on investigation are found to have resulted from some significant departures from current standards of performance.

In setting limits for rates the manufacturer has one point of view and the purchaser another. The manufacturer wishes to detect lack of control as early as possible and is willing to follow up false scents occasionally in his endeavor to prevent the persistence of costly irregularities. The purchaser is more interested in major swings or trends in quality, is not so much concerned with the use of limits for actual control and hence does not desire to instigate fruitless investigations frequently. For many telephone products, experience has indicated that a probability value between .90 and .95 is economical for shop control work while higher values such as .99 or above are better suited for quality reports issued for purposes of general information.

Inasmuch as the rate measures overall quality as determined by a number of different characteristics, its control feature relates particularly to final or partial assemblies of product. This control work should, of course, be preceded by control activities based on the same principles applied to the process inspection data for each of the essential characteristics of the parts which make up the whole.

PARTIAL SUMMARY OF INDIVIDUAL DEFECTS

Type of Defect	Demerit Weight	Base Period 1922	1923	1924	1925	1926 Jan.	Feb.	Mar.	Dec.	1926 Total	Current Year 1927 Jan.	Feb.	Mar.	Aug.
Electrical														
Type 1	100	0	0	0	1	0	0	0	16	24	18	5	14	0
Type 2	100	30	147	168	135	3	20	12	9	171	0	19	0	1
Mechanical														
Type 3	100	1	0	0	0	0	0	0	0	0	0	0	0	5
Type 4	100	0	0	9	9	0	3	1	0	10	1	1	0	0
Mounting and Assembly														
Type 5	35	5	0	1	1	2	11	0	1	44	0	0	0	0
Type 6	35	0	0	0	1	0	1	0	0	7	0	0	0	0
Type 7	20	6	11	8	8	0	0	0	1	5	0	0	1	1
Type 8	20	2	0	21	5	0	0	0	1	13	0	0	0	0
Wiring and Soldering														
Type 9	50	0	0	0	6	0	0	0	0	0	0	0	0	0
Type 10	35	0	0	6	1	0	0	0	0	0	0	0	0	0
Marking														
Type 11	50	3	4	8	8	0	0	0	1	5	0	0	0	2
Packing														
Type 12	50	13	0	0	0	0	0	0	0	0	0	0	0	0
Summary of Total Defects by Classes														
Class A	100	44	162	193	176	3	24	16	27	255	19	28	33	27
Class B	60	21	29	42	21	0	0	5	0	57	1	0	0	9
Class C	25	12	25	55	31	7	25	6	1	131	2	0	2	12
Class D	5	14	36	59	16	1	7	1	6	48	0	0	2	1
No. Inspected*		7,478	24,871	24,524	23,093	1,230	2,184	3,201	3,160	31,385	2,579	2,657	3,424	2,475
Rate		+1.05	+1.64	−0.63	+0.52	+4.9	−5.5	+2.9	+0.3	−1.29	+1.4	−1.7	−0.9	−5.9

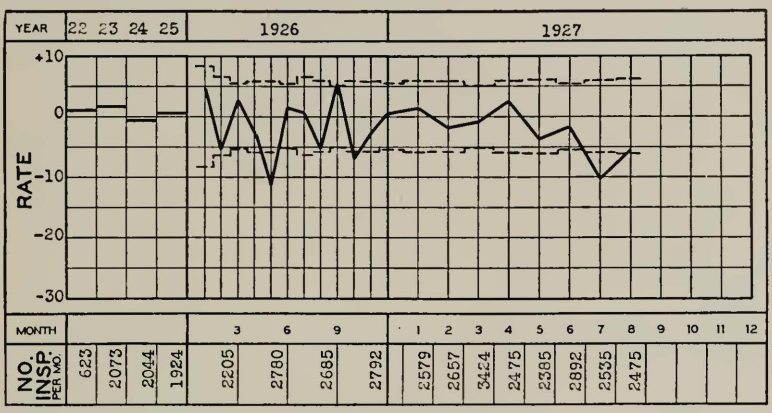
* Same for all types of defects.

Fig. 5—A typical classified summary of the defects found in inspection

Illustrative Examples

Example I—Monthly Rate for General Information Purposes Showing Quality of Finished Product

This example presents a monthly quality report for a kind of telephone apparatus which is manufactured in large quantities for use in



Defect	Wt. w	Base Period			March 1927		
		No. of Defects d	No. Insp. n	Demerits per Unit $\left(\frac{D}{n}\right)=\frac{wd}{n}$	Expected Dem. per Unit $\left(\frac{D}{n}\right)_e$	No. Insp. n	$\frac{w}{n}\left(\frac{D}{n}\right)_e$
Class A.....	100	830	111,351	.7454	.7454	3,424	.02177
Class B.....	60	170	111,351	.0916	.0916	3.424	.00160
Class C.....	25	254	111,351	.0570	.0570	3,424	.00042
Class D.....	5	173	111,351	.0078	.0078	3,424	.00001
Total				Σ.9018			Σ.02380

Computation of Control Limits.—

$$k = \frac{1}{\text{Base Period Demerits per Unit}} = 1.109, \qquad R_e = 0, \qquad K = 3,$$

$$\sigma_{R_e} = 10k\sqrt{\Sigma\left[\frac{w}{n}\left(\frac{D}{n}\right)_e\right]} = 10(1.109)\sqrt{.02380} = 1.711.$$

Limits $R_L = R_e \pm K\sigma_{R_e}$.
$= 0 \pm 3(1.711)$
$= +5.133$ and -5.133.

Fig. 6—A typical rating chart with variable control limits

central office exchanges. The data given in Fig. 5 represent a portion of the summarized results of inspection. The detailed information

given in tabulations such as this is the basic material used in investigation work relating to control of quality. The composite summary at the bottom of the figure is used directly for computing rates.

The quality report for this product is based on the following:

(1) The data are obtained by the check inspection of representative samples of the total product.
(2) The average quality for the base period is taken as the "standard expected" quality for current product. Control lines are thus placed above and below the base period rate of 0.
(3) The limits are computed each month on the basis of the number inspected during that month, using a probability value of .997.

The computations shown below the rating chart of Fig. 6 indicate the work necessary to the determination of the control limits for a given month. The basis of these computations is given in the Appendix. The limit lines on the chart are broken lines merely because the number inspected varies from month to month.

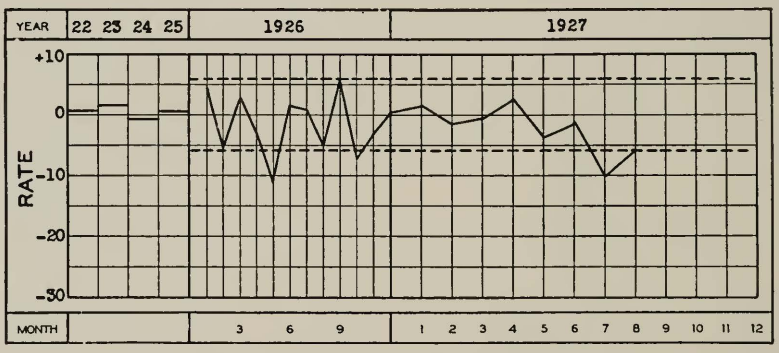


Fig. 7—Rating chart with constant control limits

When production and inspection schedules are fairly uniform so that the number inspected per month remains substantially constant, the limits may be computed once a year using some estimated size of monthly sample and drawn as parallel lines across the chart as in Fig. 7. This approximation is justified when the loss of accuracy thereby introduced is outweighed by the extra charting costs associated with monthly computations of limits.

Example II—Monthly Rates Showing Quality of a Product Before and After a Screening Inspection

Assume the following procedure to be in force.

(1) The shop product is inspected 100 per cent by an inspection group which serves as a screening medium for eliminating defective units.

(2) The product which passes this inspection is subsequently examined on a sampling basis by a check inspector who looks for the same defects.

The data obtained by the screening inspection provide a measure of the quality submitted by the Operating Department. The check inspection data give a picture of the quality of the finished product placed in stock. The rating chart is shown in Fig. 8. In a case of this sort where identical product is handled by two successive inspection groups, it is advantageous to show both rates on the same scale. In this particular instance a rate of 0 corresponds to the base period

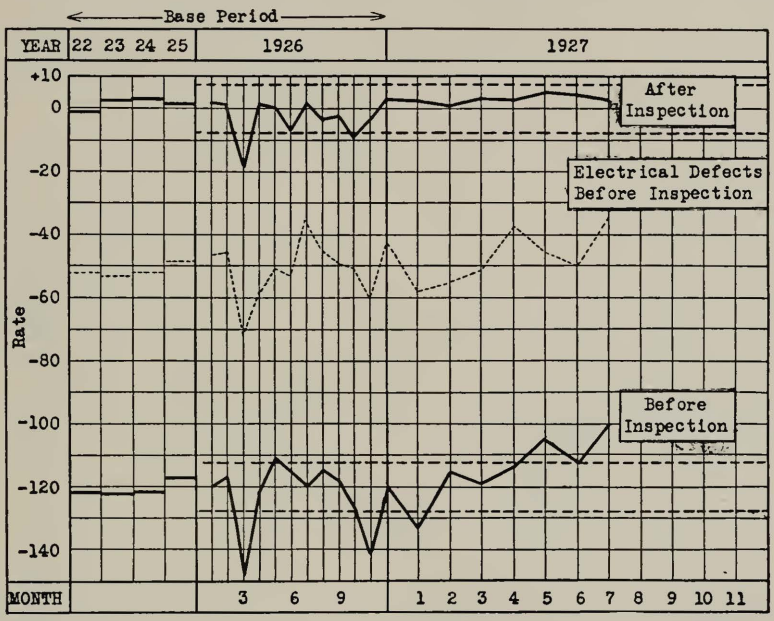


Fig. 8—Rate showing quality of a product before and after a screening inspection

quality of the product submitted to the check inspector. The control limits for the lower rate are drawn above and below the expected level of quality for product submitted to the first group of inspectors and both sets of limits are based on a probability of .95.

The results of the screening inspection can be used directly for controlling the work of the Operating Department. For this purpose it has been found valuable to prepare weekly rates with control limits based on a slightly lower probability value than that used for monthly rates. When the defects can be readily classified into two or more major groups, such as defects for electrical requirements, defects for

mechanical requirements, etc., it has often been found useful to compute sub-rates with respect to such classifications of trouble. A typical sub-rate is indicated by the fine dotted line of Fig. 9. Ex-

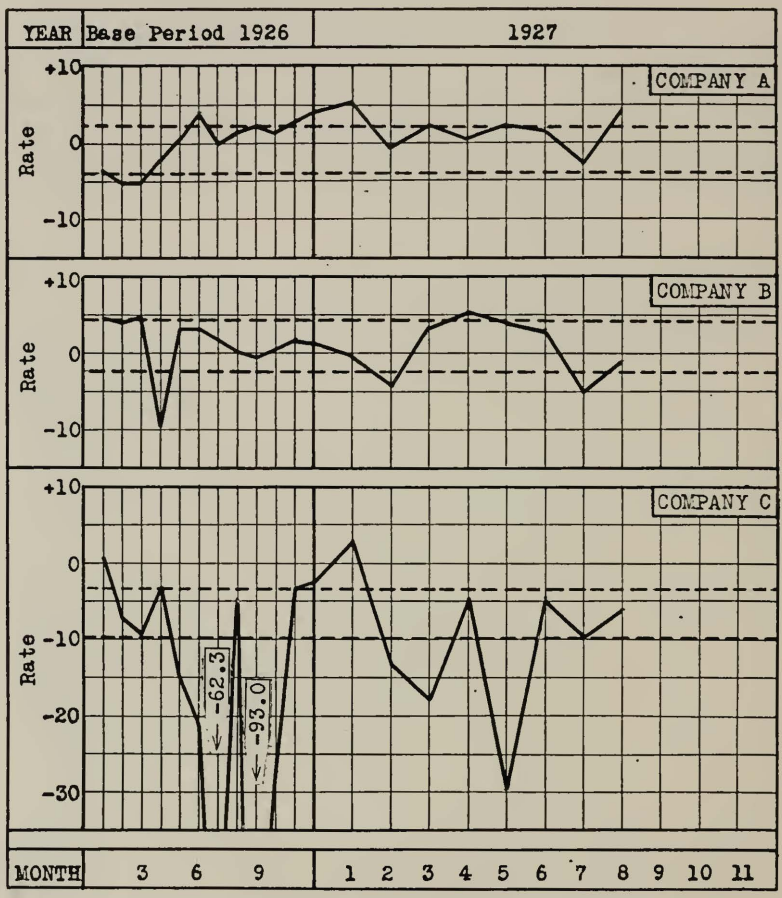


Fig. 9—Rates showing the quality of similar product supplied by three independent companies

perience has also shown the practicability of rating the quality of output of the individual operators and gangs in some classes of work for the purpose of comparing workmanship.

The principal advantage of these latter steps is to provide a ready means for tracing causes of trouble.

Example III—Rates for Comparing the Quality of Similar Product Supplied by Different Organizations

The rate can be used in like manner for comparing the quality of similar product produced by different factories of one company or by different companies when the several manufacturing units are governed by the same set of specifications.

As an example take the case of similar material supplied by three independent companies. There are several factors to be considered in constructing the rates. First of all, some standard of reference must be chosen to represent zero on the rate scale. This might correspond to the average performance of all companies, of the best company, the average of the two best, etc. Secondly, to show the degree of control for each company, the limits should be constructed independently for each company above and below the individual expectancy levels of quality.

The rating chart of Fig. 9 gives a quality picture for a particular kind of material supplied by three independent concerns, two of which are doing consistently better work than the third. Zero rate has been chosen arbitrarily to represent the average base period performance of the two best companies. The rate obtained monthly for any one company thus reflects, by its numerical magnitude, the relation between this company's current quality and that chosen as the standard of reference.

Graphical quality reports of this sort are of value to purchasing organizations in their relations with competing suppliers of similar materials.

Appendix

Computation of Control Limits

Assume

(1) $R = 0$ corresponds to base period demerits per unit.

(2) Control limits are to be set above and below some rate figure, R_e, corresponding to expected demerits per unit. (If expected quality is the same as base period quality, then $R_e = 0$.)

(3) Control limits are to be computed for monthly rates. (The procedure is similar for any other period of time.)

The following symbols will be used:

w = weight (demerits per defect) for a given type of defect.

d = expected number of defects per month, for a given type.

$D = wd$ = demerits for d defects.

n = number inspected for a given type of defect during the month.

$\left(\frac{D}{n}\right)_e$ = expected demerits per unit.

I_e = index for Expected Quality = $\frac{\text{Expected Demerits per Unit}}{\text{Base Period Demerits per Unit}}$.

R_e = rate for Expected Quality.

R_L = control limit value of rate.

$k = \frac{1}{\text{Base Period Demerits per Unit}}$.

σ = standard (root mean square) deviation.

The rate for Expected Quality is

$$R_e = 10(1 - I_e). \tag{1}$$

To find the control limits for the rate, first determine its standard deviation, σ_{R_e}.

$$\sigma_{I_e} = 10\sigma_{I_e}. \tag{2}$$

The index for Expected Quality is given by

$$I_e = k\left(\frac{w_1 d_1}{n_1} + \frac{w_2 d_2}{n_2} + \text{etc. for all types of defects}\right), \tag{3}$$

where d_1, d_2, etc. = expected number of defects per month for defects of type 1, 2, etc., and the subscripts 1, 2, etc., refer generally to the several types of defects.[7]

To find σ_{I_e}, the standard deviation of the index, the d's are considered as independent variables subject to sampling variations. I_e is then a linear function of d_1, d_2, etc., with the constant coefficients $\frac{kw_1}{n_1}$, $\frac{kw_2}{n_2}$, etc. Hence

$$\sigma_{I_e} = \sqrt{\left(\frac{kw_1}{n_1}\right)^2 \sigma_{d_1}{}^2 + \left(\frac{kw_2}{n_2}\right)^2 \sigma_{d_2}{}^2 + \cdots}. \tag{4}$$

The values of σ_{d_1}, σ_{d_2}, etc., are evaluated by the following consideration.

Assume that a sample of size N is drawn from a source for which the probability of occurrence of a defect is p. The expected number of defects in the sample is pN, and the standard deviation of the expected number is $\sqrt{p(1-p)N}$. If p is small, the factor $(1 - p)$

[7] In carrying out the computations of rates and control limits, it is convenient to group together all defects having the same "weight" (w) and the same "number inspected" (n), and to let the subscripts 1, 2, etc., of the equations refer to these groups.

can be neglected,[8] which gives $\sqrt{pN}$. Considering the practical case, if the ratio of the number of defects is small compared with the possible number of defects, then the standard deviation of the expected number, d, is equal to $\sqrt{d}$. For many telephone products certain types of defects may occur several times on a single unit of product; for example, when inspection is made for the tension requirement of springs on a relay or for character of soldered connections on a switchboard, then N refers to the number of springs or the number of soldered connections, respectively, inspected during the month. Likewise p refers to the probability of occurrence of a defective spring or of a defective soldered connection. Fortunately, in the determination of the standard deviation it is not necessary to know exactly what p is nor what N is, so long as it is known that p is small (less than .10 for ordinary engineering purposes). This condition is usually satisfied in practice, hence the above result can be used.

Equation (4) then becomes

$$\sigma_{I_e} = \sqrt{\frac{k^2 w_1^2 d_1}{n_1^2} + \frac{k^2 w_2^2 d_2}{n_2^2} + \cdots}. \tag{5}$$

To simplify routine computations, this can be changed in form by removing the factor $\frac{wd}{n}$ from each term (this is merely the Expected Demerits per Unit $\left(\frac{D}{n}\right)_e$ for a given type of defect) which gives

$$\sigma_{I_e} = k\sqrt{\frac{w_1}{n_1}\left(\frac{D}{n}\right)_{e_1} + \frac{w_2}{n_2}\left(\frac{D}{n}\right)_{e_2} + \cdots}. \tag{6}$$

and the $\left(\frac{D}{n}\right)_e$ factors may be computed directly from the totality of data available for establishing the Expected Demerits per unit.

In shorter notation, equation (6) can be expressed as

$$\sigma_{I_e} = k\sqrt{\Sigma\left[\frac{w}{n}\left(\frac{D}{n}\right)_e\right]}. \tag{7}$$

If the number of units inspected is the same for all types of defects, i.e. $n_1 = n_2 = \text{etc.} = n$, equation (7) becomes

$$\sigma_{I_e} = k\sqrt{\frac{1}{n}}\sqrt{\Sigma\left[w\left(\frac{D}{n}\right)_e\right]}. \tag{8}$$

[8] This follows directly from the Law of Small Numbers. Theoretically this result is obtained if p is small, N infinite and pN finite. See any standard text on the subject. Practically this law can be used as an approximation if p is less than .10 and N is greater than 16.

The control limits for the rate are obtained from the equation

$$R_L = R_e \pm K\sigma_{R_e}.$$

(assuming the Normal Law to be a satisfactory approximation for determining probabilities), where σ_{R_e} is given by equations (2) and (6), and where K is a constant whose value depends on the choice of probability.

The following table gives values of K for different values of probability.

Probability	K
.997	3.00
.990	2.33
.955	2.00
.900	1.65
.800	1.28
.683	1.00
.500	.675

Abstracts of Bell System Technical Papers Not Appearing in this Journal

A Note on the Thermionic Work Function of Tungsten.[1] C. DAVISSON and L. H. GERMER. It has been pointed out that the authors in a previous paper [2] had neglected to apply a correction for the "Schottky Effect" to their results. The present note applies this correction which is found to be comparatively small and does not materially affect the results given before. A fuller discussion of the interpretation of the work function measurements previously reported is also included in this note.

The Action of Fluxes in Soft Soldering and a New Class of Fluxes for Soft Soldering.[3] R. S. DEAN and R. V. WILSON. This is a report of basic studies of soldering fluxes undertaken by the authors. It was found that the fluxing action depends on the evolution at soldering temperatures of HCl gas or other halogen acid gases which have been found to be effective soldering fluxes. Based on this discovery soldering fluxes have been found among the organic compounds and the way opened for the development of a truly non-corrosive flux.

Certain Topics in Telegraph Transmission Theory.[4] H. NYQUIST. The author gives results of theoretical studies of telegraph systems which have been made from time to time. Among other things he points out that although the usual method of determining the distortion of telegraph signals is to calculate the transients of the system an alternative method is based on the steady state characteristics. For the first method the telegraph wave is taken as a function of t and for the second as a function of ω. A discussion of the minimum frequency range required for transmission at a given signaling speed is also included in the paper.

Experiments and Observations Concerning the Ionized Regions of the Atmosphere.[5] R. A. HEISING. Experiments are described in which a virtual height of the reflecting ionized region was measured using time lag between radio signals arriving over a direct and the reflected path. Heights were found of 150 to 400 miles with vertical move-

[1] *Phys. Rev.*, Vol. 30, pp. 634–638, Nov. 1927.
[2] Davisson and Germer, *Phys. Rev.*, Vol. 20, 300 (1922).
[3] *Indus. and Eng. Chemistry*, Vol. 19, No. 12, pp. 1312–1314.
[4] Presented, A. I. E. E., February 13–17, 1928. *A. I. E. E. Jour.*, Vol. XLVII, No. 3, pp. 214–216, Mar. 1928.
[5] *Proc. I. R. E.*, Vol. 16, pp. 75–99, Jan. 1928.

ments at rates as high as 20 miles per minute. Other experiments and curves are mentioned which show absorption to be one of the important factors causing poor daylight transmission for wave-lengths around 214 meters. A discussion is given to show that both electromagnetic waves from the sun and β particles must be assumed to produce ionization to explain radio transmission phenomena observed.

The ionization is pictured as beginning at an altitude of about 16 miles and extending upward, and as experiencing diurnal and seasonal variations. The electromagnetic or day ionization occupies a wide region, and is fairly steady except for the diurnal variation. The β particle ionization which is the principal ionization at night occurs continuously. It is, however, less dense than the other ionization and is very variable.

Diffraction of Electrons by a Crystal of Nickel.[6] C. DAVISSON and L. H. GERMER. The scattering of electrons by nickel has been reported on recurrently since 1921. This paper gives the latest results which indicate that a wave-length is in some way connected with the electron's behavior. A rather complete summary of the experiments and conclusions appeared in the *Bell System Technical Journal* for January, 1928, which makes unnecessary a fuller account here.

A General Operational Analysis.[7] W. O. PENNELL. The paper outlines the elements of a general operational analysis. Using p as the symbol for an operator, the author, starting with a general typical defining equation such as $p(ax^b) = \psi(a, x, b)$, proceeds by definitions and theorems to demonstrate the use of operational methods in a large variety of common mathematical operations.

Precision Determination of Frequency.[8] J. W. HORTON and W. A. MARRISON. The relations between frequency and time are such that it is desirable to refer them to a common standard. Reference standards, both of time and of frequency, are characterized by the requirement that their rates shall be so constant that the total number of variations executed in a time of known duration may be taken as a measure of the rate over shorter intervals of time. Frequency standards have the further requirement that the form of their variations and the order of magnitude of their rates shall be suitable for comparison with the waves used in electrical communication.

Two different types of standard which meet these requirements are

[6] *Phys. Rev.*, Vol. 30, No. 6, pp. 705–740, Dec. 1927.
[7] *Jl. of Math. and Phys. of M. I. T.*, Vol. 7, No. 1, pp. 24–38, Nov. 1927.
[8] *Proceedings of the I. R. E.*, Vol. 16, No. 2, Feb. 1928.

described. One consists of a regenerative vacuum-tube circuit, the frequency of which is determined by the mechanical properties of a tuning fork. The other is a regenerative circuit controlled by a piezo-active crystal. Means are provided, in the case of each standard, whereby the recurrent cycles may be counted by a mechanism having the form of a clock, the rate of which is a measure of the frequency of the reference standard.

Data taken over a period of several years with a fork-controlled circuit show that, under normal conditions, its rate may be relied upon to two parts in one million. Data taken over a much shorter time with crystal controlled oscillators indicate that they are about ten times as stable.

Plane Waves of Light. II. Reflection and Refraction.[9] THORNTON C. FRY. This paper extends the study of plane light waves, which was begun in the *Journal of the Optical Society* for September, 1927, to the phenomena of reflection and refraction. It develops the general formulæ for reflection from a plane boundary between two media and for reflection from thin films, paying especial attention to the situations under which "hybrid" polarization occurs. The differences between the reflected and refracted components in the case of dielectrics and metals are illustrated by a number of diagrams.

The paper closes with a discussion of the determination of the optical constants of metals, both from the state of polarization of the reflected light and from the direction of emergence of the ray transmitted through a prism.

Propagation Characteristics of Sound Tubes and Acoustic Filters.[10] W. P. MASON. This paper describes a method for making acoustic propagation measurements, and presents the results of attenuation and velocity measurements of straight tubes and acoustic filters. The ordinary electrical transmission measuring circuit is employed in conjunction with loud speakers and acoustic resistance terminations. The process of measurement consists in obtaining the transmission characteristics of the systems with the device to be measured in the acoustic circuit, then obtaining the characteristics with the device to be measured taken out of the acoustic circuit. The difference between these two measurements gives the transmission characteristics of the device to be measured between the two acoustic resistance terminations. Results obtained from these measurements for straight pipes agree well with the Helmholtz-Kirchoff Law.

[9] *Journal of the Optical Society of America*, Vol. 16, pp. 1–25, January, 1928.
[10] *Physical Review*, pp. 283–295, Vol. 31, No. 2, February, 1928.

Brittleness Tests for Rubber and Gutta-Percha Compounds.[11] G. T. KOHMAN and R. L. PEEK, JR. An insulating material compounded of rubber, gutta-percha, or of similar substances becomes brittle at a temperature, characteristic of the material, below which it may not be used if liable to mechanical stress. This paper describes an apparatus designed to determine this temperature by giving the sample a sharp bend through a fixed angle. The highest temperature at which fracture occurs in this test (the brittle temperature) is found to be nearly independent of the bending angle and the sample's dimensions provided the rate of bending is maintained at a nearly constant (high) rate. A modified form of the apparatus is also described with which the brittle temperature may be determined when the material is under high hydrostatic pressure. The constancy of the brittle temperature when determined under different conditions suggests that it marks a change in the structure of the material.

[11] *Ind. and Eng. Chem.*, Vol. 20, pp. 81–83, January, 1928.

Contributors to this Issue

O. B. Blackwell, B.S. in electrical engineering, Massachusetts Institute of Technology. After graduation, he entered the Engineering Department of the American Telephone and Telegraph Company as engineer and in 1919 was made Transmission Development Engineer. Mr. Blackwell has general supervision of transmission developments and by virtue of his position has been prominently associated with progress in long distance wire and radio telephony.

K. W. Waterson, S.B. in E.E., Massachusetts Institute of Technology, 1898; Mechanical Department, American Bell Telephone Company, 1898; in charge of equipment engineering, 1901; in charge of traffic engineering, 1905; in charge of traffic and equipment engineering, 1906; Assistant Chief Engineer, 1907; Engineer of Traffic, 1909; Executive Officer, Department of Development and Research, 1919; Assistant Chief Engineer, Department of Operation and Engineering, 1920; Assistant Vice President, 1927, in charge of traffic, plant operation and general results divisions, Department of Operation and Engineering.

Sallie Pero Mead, A.B., Barnard College, 1913; M.A., Columbia University, 1914; American Telephone and Telegraph Company, Engineering Department, 1915–19; Department of Development and Research, 1919–. Mrs. Mead's work has been of a mathematical character relating to telephone transmission.

Oliver E. Buckley, B.Sc., Grinnell College, 1909; Ph.D., Cornell University, 1914; Engineering Department, Western Electric Company, 1914–17; U. S. Army Signal Corps, 1917–18; Engineering Department, Western Electric Company (Bell Telephone Laboratories), 1918–. During the war Major Buckley had charge of the research section of the Division of Research and Inspection of the Signal Corps, A. E. F. His early work in the Laboratories was concerned principally with the production and measurement of high vacua and with the development of vacuum tubes. More recently he has directed investigations of magnetic materials and the development of the permalloy-loaded submarine cable.

JOHN R. CARSON, B.S., Princeton, 1907; E.E., 1909; M.S., 1912; Research Department, Westinghouse Electric and Manufacturing Company, 1910–12; instructor of physics and electrical engineering, Princeton, 1912–14; American Telephone and Telegraph Company, Engineering Department, 1914–15; Patent Department, 1916–17; Engineering Department, 1918; Department of Development and Research, 1919–. Mr. Carson's work has been along theoretical lines and he has published several papers on theory of electric circuits and electric wave propagation.

KARL K. DARROW, S.B., University of Chicago, 1911, University of Paris, 1911–12, University of Berlin, 1912; Ph.D. in physics and mathematics, University of Chicago, 1917; Engineering Department, Western Electric Company, 1917–24; Bell Telephone Laboratories, Inc., 1925–. Mr. Darrow has been engaged largely in preparing studies and analyses of published research in various fields of physics.

C. D. HART, M.E., Cornell University, 1906; entered Western Electric Company in Student Course at New York in 1906; transferred to Hawthorne in 1911, development work on the manufacture of lead-covered cable; transferred to Tokyo, Japan, in 1913 to inaugurate the manufacture of lead-covered telephone cable at the Nippon Electric Company; returned to Hawthorne, December, 1915; 1916–20, general foreman of Cable Shops, Metal Finishing Department and Rubber Shops; 1920–23, manufacturing development work; 1923–, Assistant Superintendent of Manufacturing Development.

J. HERMAN, E.E., Lehigh University, 1920; Department of Development and Research, American Telephone and Telegraph Company, 1920–. Mr. Herman has been engaged chiefly in telegraph transmission development work and has been associated with the development of voice-operated switching devices.

H. F. DODGE, S.B., Mass. Inst. Tech., 1916; instructor in electrical engineering, 1916–17; A.M., Columbia University, 1922; Engineering Department of the Western Electric Company and Bell Telephone Laboratories, 1917–. Mr. Dodge was earlier associated with the development of telephone instruments and allied devices, and is now engaged in development work relating to the application of statistical methods to inspection engineering.

The Bell System Technical Journal

July, 1928

Precision Tool Making for the Manufacture of Telephone Apparatus

By J. H. KASLEY and F. P. HUTCHISON

THERE is probably no field of human endeavor in which hand labor has been more completely replaced by labor-saving devices than in the field of manufacturing. The design and employment of special tools together with semi- and full automatic machinery for

ERRATA: *Bell System Technical Journal*, April, 1928

Page 327, Table 2—Interchange the number "200" of column 6 and number "600" in column 8.

Page 328, beginning line 4, should read—(*a*) Phantom to phantom; 1 represents the two wires, connected in parallel, of one pair of a quad. 2 represents the two wires in parallel of the other pair of the quad, and 3 and 4 represent similarly the pairs of another quad.

Page 347—Figure 3 should be inverted.

tool making art as practiced by this Company and to do this, illustrative material will be drawn from among the large number of punches and dies used for punch press methods of manufacture. The methods employed and precision necessary in building the tools discussed below can be considered as representative of the high class of workmanship required throughout the Company's tool rooms.

Punches and Dies

Tool Making for Telephone Apparatus Manufacture. Briefly, a punch and die comprises a pair of individual tools so constructed

John R. Carson, B.S., Princeton, 1907; E.E., 1909; M.S., 1912; Research Department, Westinghouse Electric and Manufacturing Company, 1910–12; instructor of physics and electrical engineering, Princeton, 1912–14; American Telephone and Telegraph Company, Engineering Department, 1914–15; Patent Department, 1916–17; Engineering Department, 1918; Department of Development and Research, 1919–. Mr. Carson's work has been along theoretical lines and he has published several papers on theory of electric circuits and electric wave propagation.

Karl K. Darrow, S.B., University of Chicago, 1911, University of Paris, 1911–12, University of Berlin, 1912; Ph.D. in physics and mathematics, University of Chicago, 1917; Engineering Department, Western Electric Company, 1917–24; Bell Telephone Laboratories, Inc.. 1925–. Mr. Darrow has been engaged largely in preparing

[illegible], S.B., Mass. Inst. Tech., 1916; instructor in electrical engineering, 1916–17; A.M., Columbia University, 1922; Engineering Department of the Western Electric Company and Bell Telephone Laboratories, 1917–. Mr. Dodge was earlier associated with the development of telephone instruments and allied devices, and is now engaged in development work relating to the application of statistical methods to inspection engineering.

The Bell System Technical Journal

July, 1928

Precision Tool Making for the Manufacture of Telephone Apparatus

By J. H. KASLEY and F. P. HUTCHISON

THERE is probably no field of human endeavor in which hand labor has been more completely replaced by labor-saving devices than in the field of manufacturing. The design and employment of special tools together with semi- and full automatic machinery for operating them have reached a high stage of development and are probably more responsible than any other factors for the present age being generally referred to as the industrial age. The notable economies of present day manufacture result no more from the rapid production of parts thus made possible than from the interchangeability of these parts because of the accuracy with which they have been produced.

At the foundation of precision manufacture by machine lies the art of tool making. As a result of the impetus given it by the economic justification underlying the transition from hand labor to mechanical devices, it has grown steadily in importance and in refinement. In large measure, it is the art of tool making which insures the interchangeability of product.

There are probably few industries in which the refinements of the tool making art have been carried further than in the manufacture of telephone apparatus and equipment, especially when handled on a large production basis as by the Western Electric Company. The purpose of this article is to outline some of the refinements of the tool making art as practiced by this Company and to do this, illustrative material will be drawn from among the large number of punches and dies used for punch press methods of manufacture. The methods employed and precision necessary in building the tools discussed below can be considered as representative of the high class of workmanship required throughout the Company's tool rooms.

Punches and Dies

Tool Making for Telephone Apparatus Manufacture. Briefly, a punch and die comprises a pair of individual tools so constructed

with respect to each other that, when properly guided and forced into engagement with sufficient pressure, they will produce a uniform permanent change on the material placed between them. Punches and dies are made to perform a variety of operations, such as cutting or shearing parts from strip stock, commonly termed blanking, perforating or piercing holes, drawing, forming or bending, stamping, embossing, etc. In many instances two or more operations are combined in one tool, as, for example, a perforating and blanking punch and die, which cuts the part to its required shape and also perforates the required holes. Multiple operation tools may be constructed in many different ways, depending on the particular requirements of the part to be made. Typical illustrations of punches and

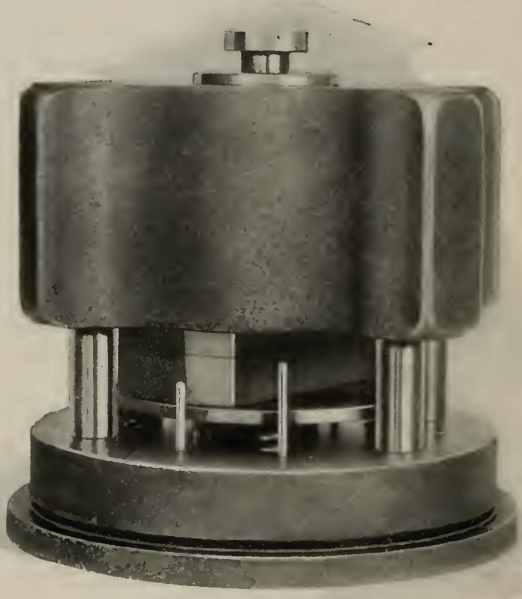


Fig. 1—Compound punch and die of the liner pin type assembled.

dies for accurate work are shown in Figs. 1 and 2. The former shows a compound punch and die of the liner or guide pin type assembled, and the latter shows a partially disassembled tool of the sub-press design, in which the moving member is completely enclosed and guided by the housing.

The compound type of construction mentioned in the preceding paragraph, which perforates and blanks the part complete in one die position and one stroke of the press, gets its name from this feature of performing a compound operation in one die position, and is generally used where very accurate parts, practically free from distortion and

with clean-cut edges, are to be produced, and particularly where thin stock is used. It is also preferable to other types of tool construction when the part is irregular in shape or is to be produced in large quantities, because of the uniformity of product, high speed at which it can be operated and because of its long life. Where small holes are to be perforated, the compound type is often advisable due to the fact that the perforators can be supported more substantially, with reduced breakage.

Fig. 3 is a cross-sectional view of one of the standard designs of sub-press compound tools illustrating this type of construction. As its name implies, the sub-press type is a practically self-contained press which is placed, assembled, in the power press. As will be noted from this figure, the compound type of tool has the perforating punches

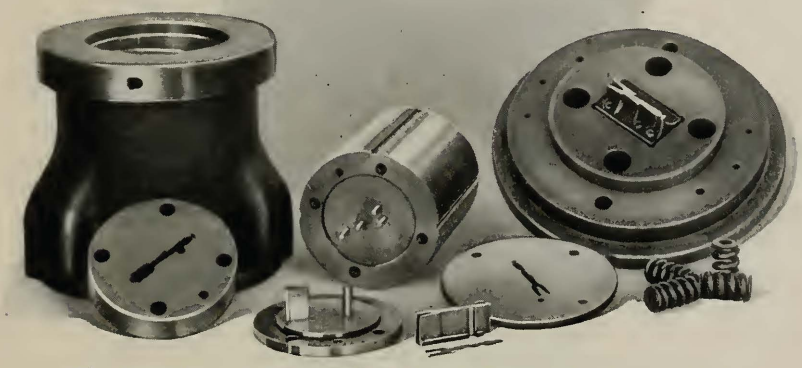


Fig. 2—Sub-press compound punch and die partially disassembled to show construction.

L located inside the blanking die *N*, and supported by the shedder *M*, and the die openings for the perforators inside punch *P*, which is fastened to the base *H* of the tool. In operation, the base *H* is mounted on the bed of the press and the cap adapter *A*, which is attached to the plunger *D*, is fastened to the slide or ram of the press. The stock is fed over the stripper *O* and the die *N* descends, thus depressing the stripper *O* and causing the shedder *M* to recede into the die *N*. As the shedder is backed up by a heavy spring *C*, the metal being blanked is held under pressure between the shedder and the punch *P* so that this type of construction fabricates thin sheet metal under conditions which insure the best results. As the downward movement progresses, the blank is cut from the stock and the holes perforated. The slugs forced out by perforators *L* drop through

the punch P, and as the die ascends on the up stroke of the press the blank is forced back into the stock by the action of the stripper and

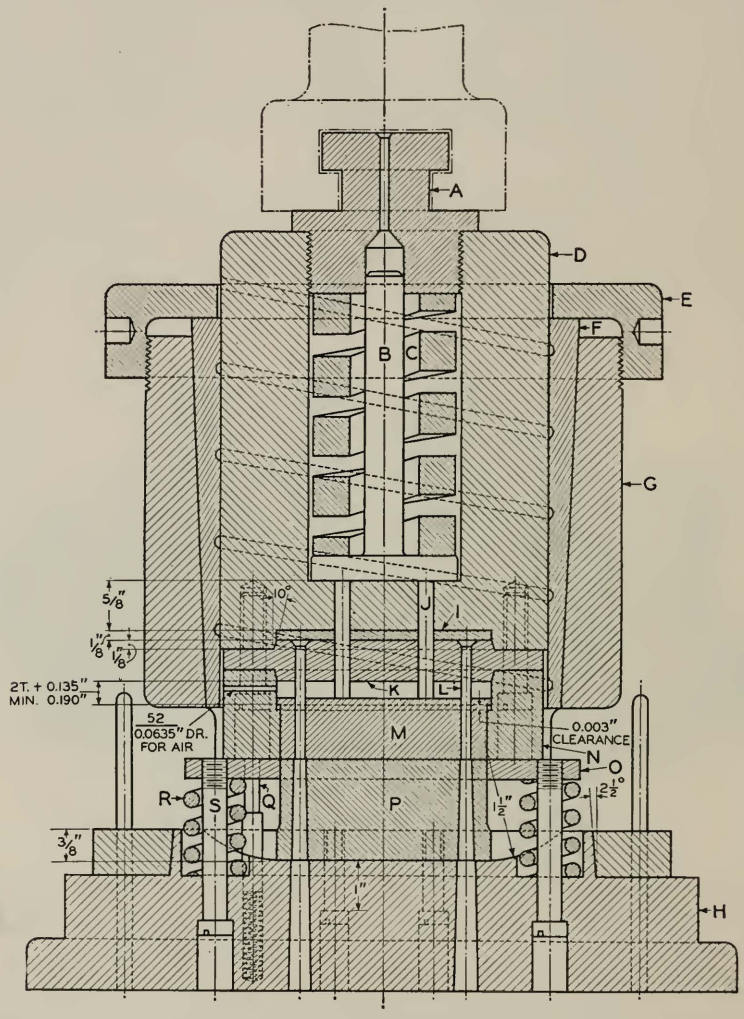


Fig. 3—Cross-sectional view of standard design for sub-press compound punch and die illustrating principal parts.

shedder. The material is then advanced to the next position and the operation repeated.

If no positive provision is provided in a punch and die for securing fixed alignment of the cutting members, considerable care and skill is required in order to adjust and set the tool in the press so as to bring

the die opening into its proper position with respect to the punch. Also, after the tool is set there is the possibility, especially in the case of the higher speed presses operating at about 300 strokes per minute, of the die shifting during operation and resulting in the "shearing" of the cutting edges of the die and punch. To overcome this difficulty the liner pin type of construction illustrated in Fig. 1, and the sub-press type shown in Figs. 2 and 3 are used in the better grade tools, and especially where a very small clearance must be maintained between the punch and die opening. In the former the arrangement consists of two or more round guide rods or liner pins fastened in the

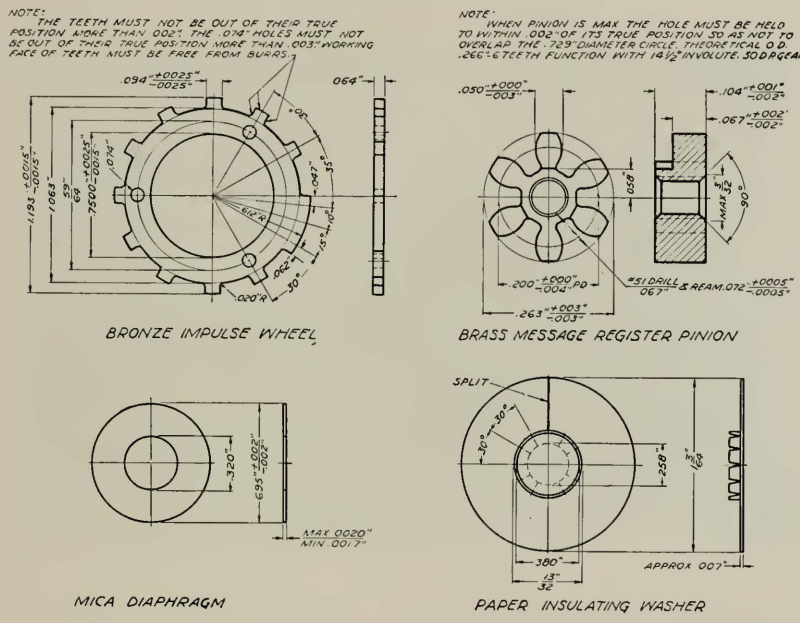


Fig. 4—Sketches of typical piece parts requiring accurately built punches and dies

bottom part of the tool and passing up through accurately bored holes in the top part, thus insuring that both members are always in their proper relative position. While the liner pin type of tool affords a very satisfactory alignment in most cases, the sub-press type is the better construction which, because of its "piston and cylinder" design, insures a more positive alignment. This is illustrated in Fig. 3. The plunger *D*, to which is attached the die *N*, perforators *L*, etc., slides in the bushing *F* in the housing *G*, and is therefore always in alignment with the base and the punch attached to it. This type of construction is followed largely where the part is small enough so

that it can be adapted to standardized housings, which are stocked, and especially where the tool is to be operated on high speed presses.

In order that a better appreciation may be obtained of some of the more exacting requirements which are being met in building tools of this kind, several of the important reasons for accurate workmanship to limits as close as a few ten-thousandths of an inch or less on some of the tool parts, together with typical examples, will first be considered.

Meeting the Accuracy Required of Piece Part. In order to obtain the correct functioning of the apparatus or equipment and also to insure interchangeability in assembly, many piece parts must be made with a

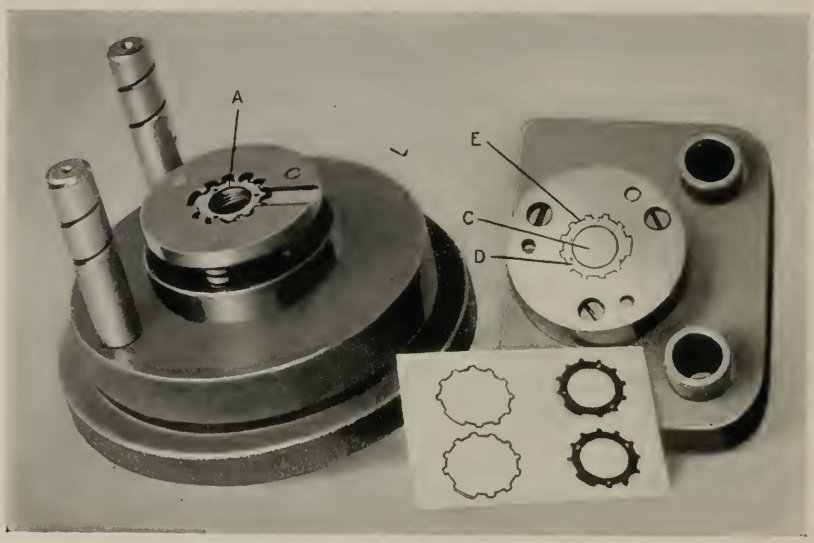


Fig. 5—Shaving and perforating punch and die for impulse wheel.

considerable degree of accuracy, often within limits of ± .001 in. or less for some of the dimensions. This is one of the most important and common reasons for accurately built tools, especially in the case of parts made in sufficiently large quantities to require a number of similar tools producing the same part, as the product of each tool must be interchangeable with that of any other provided for the same operation, and also for subsequent operations.

The bronze impulse wheel for No. 2 type dials and the pinion used in message registers, Fig. 4, are typical examples. Both of these parts are given shaving operations—the wheel after being blanked, and the pinion after being cut to length and swaged—in order to secure the required accuracy and smoothness of contour. Fig. 5

shows the shaving and perforating punch and die for the impulse wheel, together with an illustration of the part and the stock removed from the outer edge in the shaving operation which amounts to about .008 in. This tool is built to have a clearance between the shaving punch *A* and the die opening *B* of only two to four ten-thou-

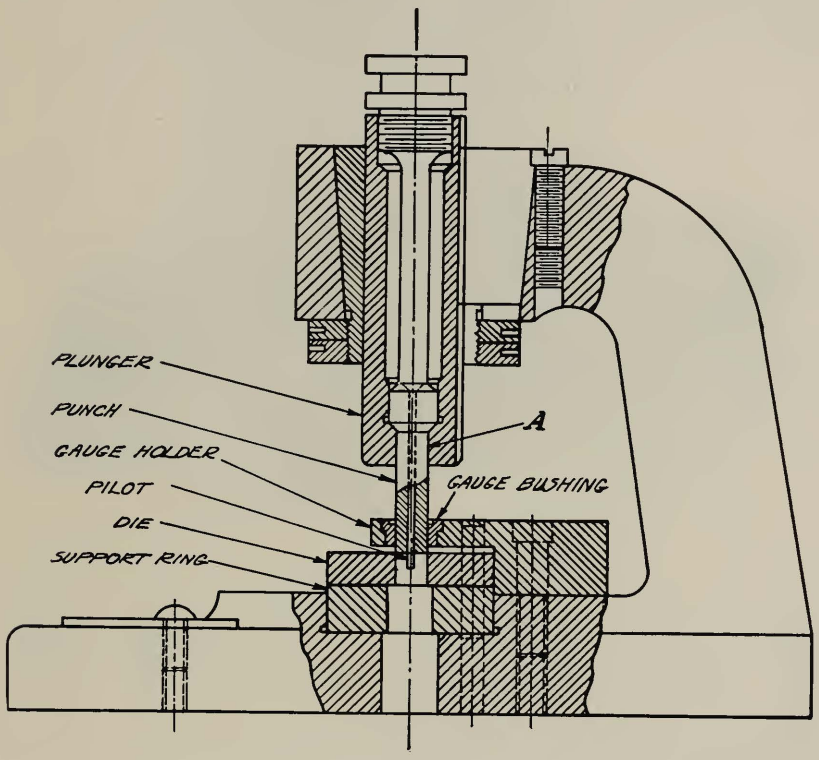


Fig. 6—Cross-section of second operation shaving punch and die for message register pinion.

sandths of an inch all around, and the punch must be located very accurately so that it will center in the die within this amount of clearance. The shaving perforator *C* for the center hole also has practically the same limit as has also the shedder *D*. Holes or openings in the die are held to as close as .0005 in. of their nominal dimensions and also with respect to each other. Fig. 6 is a partial cross-section of the second operation shaving punch and die for the message register pinion, which is made to similarly accurate limits. However, the close workmanship on parts of this tool is also necessary in order to insure interchangeability of tool parts, which will be referred to later.

One of the best examples of a high grade tool needed to make parts within close limits is the compound punch and die shown partially completed and disassembled in Fig. 7, for perforating and blanking complete in one operation the multiple bank terminal strip shown in front of the tool in the illustration. This terminal strip is $36\frac{3}{4}$ in. long and has 30 common terminals on each side spaced on $1\frac{1}{4}$ in. centers and 129 perforated holes. The design of the part requires that all terminals and some of the holes be held to within limits of $\pm$ .004 in.

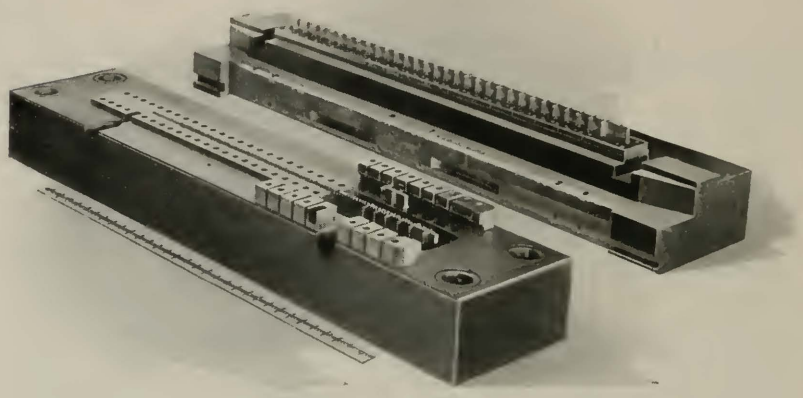


Fig. 7—Compound punch and die for blanking and perforating multiple bank terminal strip.

of their correct location with respect to a designated hole at one end of the strip. A limit of $\pm$ .004 in. for a single dimension is ordinarily not a difficult one to work to, but in this case the fact that any inaccuracies are accumulative makes it a very difficult limit to meet.

To make this tool with the required limits of accuracy, it is necessary to make the punch, die, shedder and punch plate sections, as shown in the illustration, so nearly to their exact dimensions that they are practically interchangeable. In fact, the variations are so small that if all the sections were removed from the tool and reassembled in different positions and combinations, the changed tool would not vary more than $\pm$ .0002 in. from the previous dimension over the entire length of the sections. In assembling the sections, it is necessary that they be carefully cleaned, as a slight amount of oil or dirt between them would throw the tool outside the desired limits. When it is considered that the tool must be made to much closer limits than the piece part, that all of the 103 sections, as well as the perforators, must

fit together within an overall limit of + .001 − .000 in., that the punches and perforators must be accurately centered in the die openings with a clearance all around for the punches of .0003 in. to

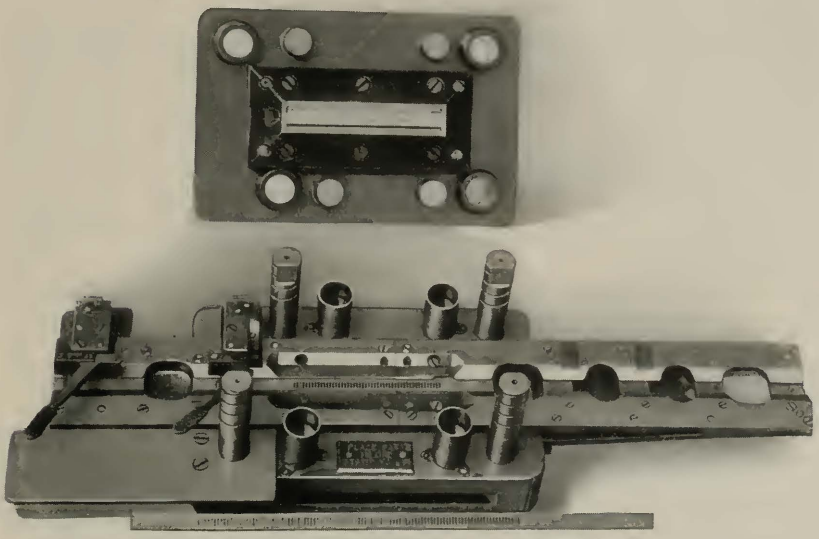


Fig. 8—Second operation punch and die for perforating holes in rack.

.0005 in. and for the perforators about .0001 in., it is apparent why, with the possibility of the errors being accumulative, each individual section must be made very accurately, the majority not exceeding a

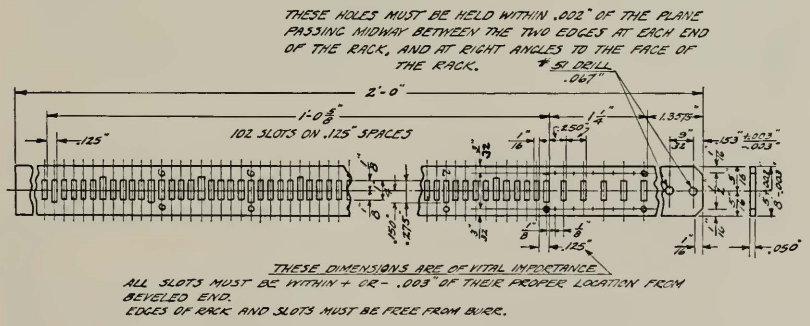


Fig. 9—Phosphor bronze rack for panel dial equipment.

limit of ± .00001 in. In addition, the slot in the die holder which holds the die, shedder, and punch plate sections, is made to almost exact dimensions and must have its sides parallel with each other and perpendicular to the bottom surface.

Another tool of this kind is the second operation perforating punch and die shown in Fig. 8, which perforates the slots in the rack for elevator apparatus shown in Fig. 9, and also in the illustration of the tool. The requirements of this part specify that all of the 107 slots 1/16 in. wide and spaced on .125 in. centers must be within ± .003 in. of their proper location from the beveled end, which is a difficult requirement to meet on account of the nature and thickness of the stock, and, like the multiple bank strip, there must be practically no accumulated error. The tool has fifty perforators, fifty-one die and fifty-two shedder sections which are made with a degree of accuracy comparable to that of the multiple bank strip tool.

Insuring Accurate Gaging in Subsequent Operations. This is a reason which sometimes requires the tool maker to work to closer limits or to hold certain dimensions to closer limits than would be otherwise required. In such cases the tool must produce piece parts which are sufficiently accurate at certain gaging points used later in other tools or in the apparatus assembly fixtures, to insure the proper results from the subsequent tools and in assembly. This may require an accuracy of from .0005 in. to .001 in. An example is the bank contact for step-by-step type banks. In order that the parts may be made sufficiently accurate at certain points so that proper bank assembly may be obtained with the assembly fixtures, the blanking die openings are made to a limit of + .001 in. − .000 in. for the width at the ends, the length, and the offset dimension, although the apparatus requirements for the piece part do not necessitate this degree of accuracy.

Production of Satisfactory Blanks from Thin Stock. Typical piece parts of this kind are the mica diaphragm .0017 in. to .002 in. thick used in transmitters, and the oiled red rope paper insulator .007 in. thick for coil spool assemblies which are shown in Fig. 4. In order to obtain clean-cut blanks, with practically no rough edges or burrs, from thin material of this kind, it is necessary that the clearance between the blanking punch and the die for the insulator does not exceed .0002 in. and the diaphragm .0001 in. all around, and the perforators nearly the same. In fact, these die parts are made to fit so closely that they will cut wet tissue paper. Accurate working fits are necessary between the other moving members, such as shedder, liner pins, etc., which mean that it must be possible to just push the parts together with no perceptible shake or clearance. The general construction of these tools is of the standard compound liner pin type similar to Fig. 1.

Interchangeability of Tool Parts. Some tools are so designed that certain parts, which on account of the design of the part being produced

are of fine construction, may be readily replaced in case of breakage or wear. In this case it is necessary that the parts be made accurately where they fit together, in order to insure interchangeability of the tool parts, without affecting the satisfactory operation of the tool or the accuracy of the parts being produced. The shaving punch and die for the message register pinion previously referred to and shown in Fig. 5 is a typical example, the construction and limits being so that parts such as the punch, die, gage bushing, and pilot may be easily replaced. For instance, in order to insure interchangeability of the punch, the dimensions of the plunger and punch at *A* are held within a limit of .0002 in., and other parts to correspondingly close limits.

Feeding of Material and Properly Formed Part in "Tandem" or "Follow" Type of Dies. In this type of tool the operation is a progressive one. While one part of the die notches, embosses, forms, or

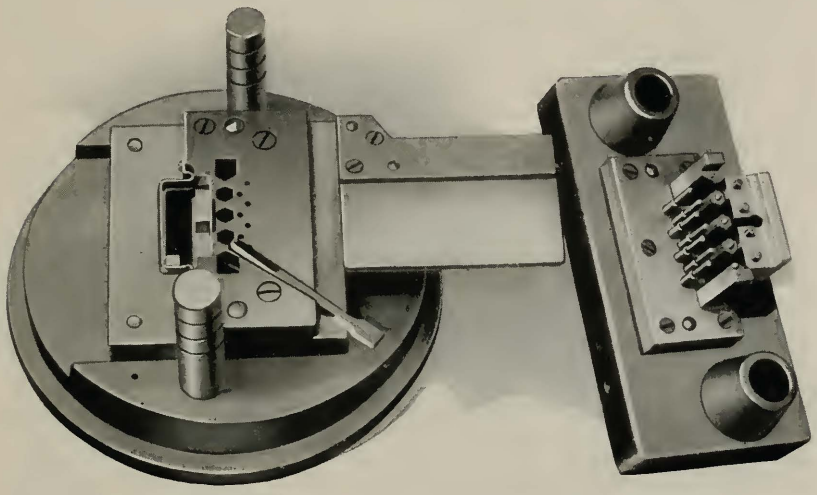


Fig. 10—Multiple perforating, blanking, and clipping punch and die for 3/8" brass hexagon nuts.

perforates the stock, another part blanks out the parts at a place where, at a former stroke, the preceding operations have been performed, so that complete parts result from each stroke of the press, although, of course, more than one operation has been performed on the parts before completion. This is illustrated in Fig. 10, which shows a multiple perforating, blanking, and clipping punch and die for making 3/8 in. brass hexagon nuts. As will be noted from the construction of the tool and the sequence of operations shown in Fig. 11, seven nuts are made with no scrap skeleton remaining at each stroke

of the press, three of the parts falling through the blanking openings and the others through the rectangular opening after being clipped off at the edge of the die.



1st Step
Perforate 7 holes and clip edges of stock



2d Step
Perforate 2d set of holes, blank out 3 nuts and notch out nuts along edges of stock



3d Step
Perforate 3d set of holes, blank out 2d set of 3 nuts, clip off 4 nuts and cut 2d notch along edge of stock

Fig. 11—Steps in the manufacture of brass hexagon nuts by scrapless punch and die method

On this tool, accuracy, from a tool making standpoint, is necessary in order to insure proper feeding of the material and an equal sided

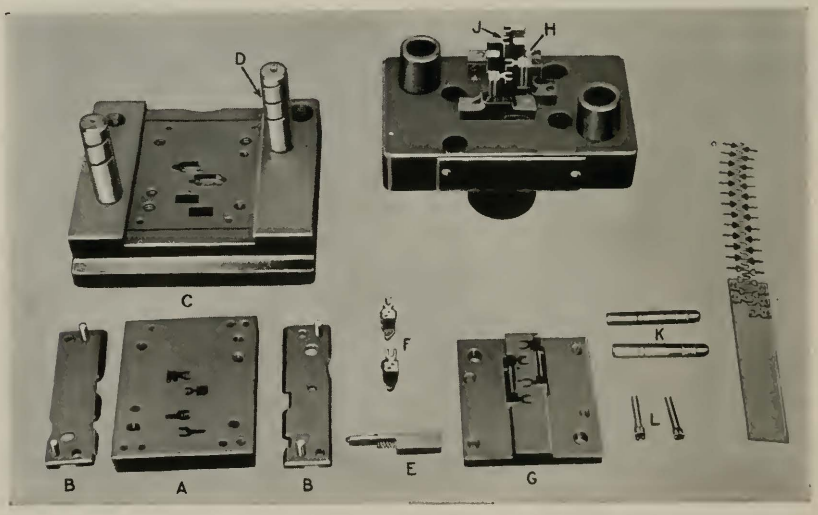


Fig. 12—Punch and die for shearing, blanking, and forming solderless cord tip

product. When the first tools were built, it was found necessary to develop very accurately the distance between the perforator and

blanking openings on account of the elongation or "creep" of the material. An error in a tool of this type is detected very easily in the product, and it requires only a very small amount to make the hexagon nut irregular in shape or "lop-sided" with respect to the center hole. It is therefore necessary that the tool maker work to close limits in maintaining the relationship between the perforator and the blanking openings and clipping edge, and the total variation between any of these is not more than .0008 in. to .001 in. It is not only necessary that this accuracy be held on the die section, but also on the punch plate and stripper, in order to insure the proper clearance between the punch members and die openings, which is .0015 in.

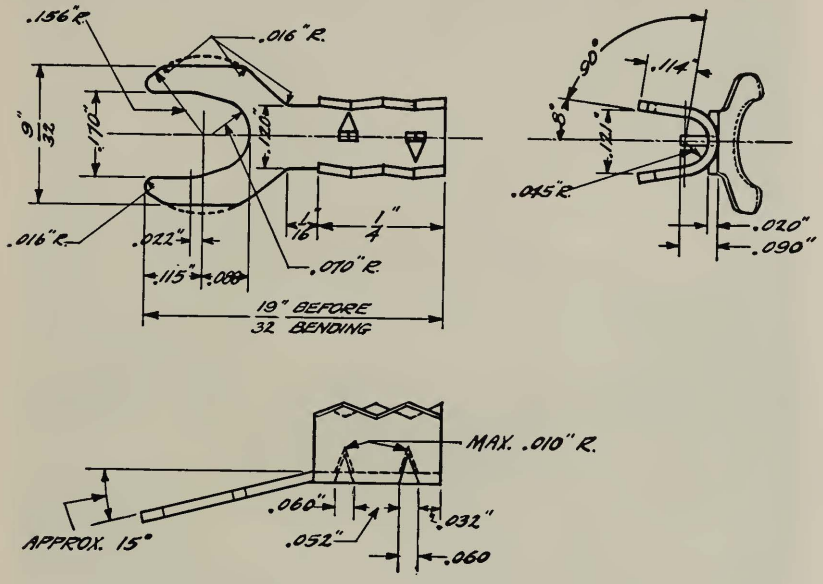


Fig. 13—No. 92 solderless cord tip.

Another example is the shearing, blanking, and forming punch and die shown partially dismantled in Fig. 12, which is used for making the solderless cord tip, Fig. 13. The parts of the tool, as shown in the figure, are *A* die, *B* stock guides, *C* die holder, *D* liner pins, *E* finger stop, *F* shedders, *G* stripper plate, *H* punch holder, *I* blanking punches, *J* forming punches, *K* dowel liner pins, *L* shearing or perforating punches. The blanked strip in the illustration shows the sequence of operations, the parts being first blanked and sheared two at a time with the blanks remaining in the scrap skeleton. The stock then advances until the blanks register with the forming die where the saw tooth

portion of the part is bent up, two complete parts being made with each stroke of the press. The location of the forming section with respect to the blanking section in this tool must be very accurate, as otherwise the blank will not register exactly under the forming punch and an incorrectly formed part will result. Also, the forming punches must be located accurately so they will center in the die openings. Other features

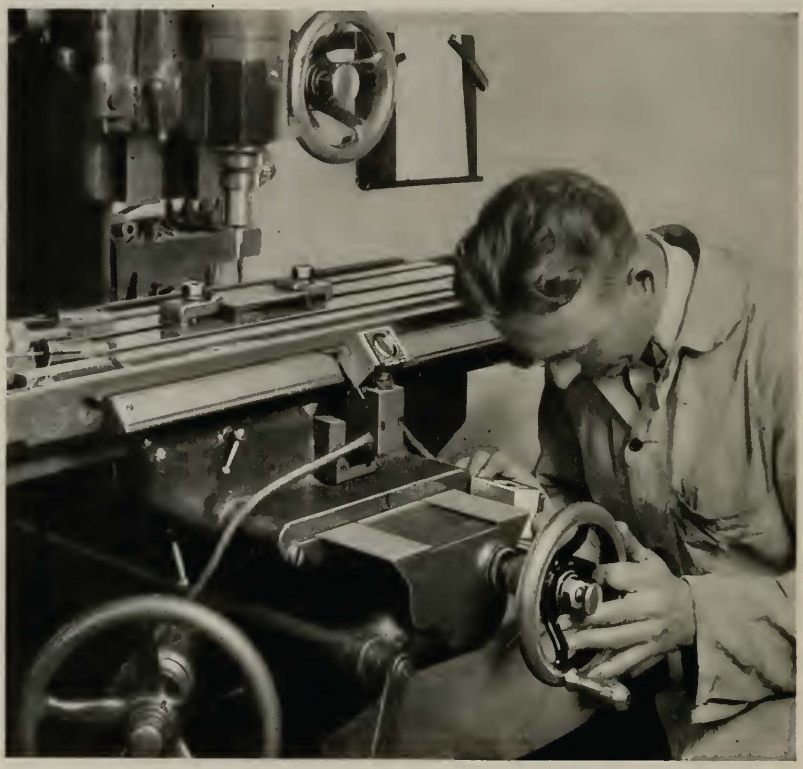


Fig. 14—Boring datum holes in master plate on veneer milling machine

regarding the workmanship required on this tool are included in the description of its construction given in the following paragraphs. Stock is fed to the punch and die by an automatic roll feed with an adjustment provided such that a precision feed within $\pm$ .0005 in. of the nominal may be obtained, which insures each part being located in the proper position for forming.

Making Punch and Die Sections

Various methods may be used in the making of punch and die sections depending on the character of the work, accuracy required, etc. The use of templates and the vernier height gage in laying out work centers and outlines, the application of the master plate method, the use of micrometer heads and verniers on milling machines and various systems of end and distance gages are all found of value in this work. A brief description of several of the operations performed in making

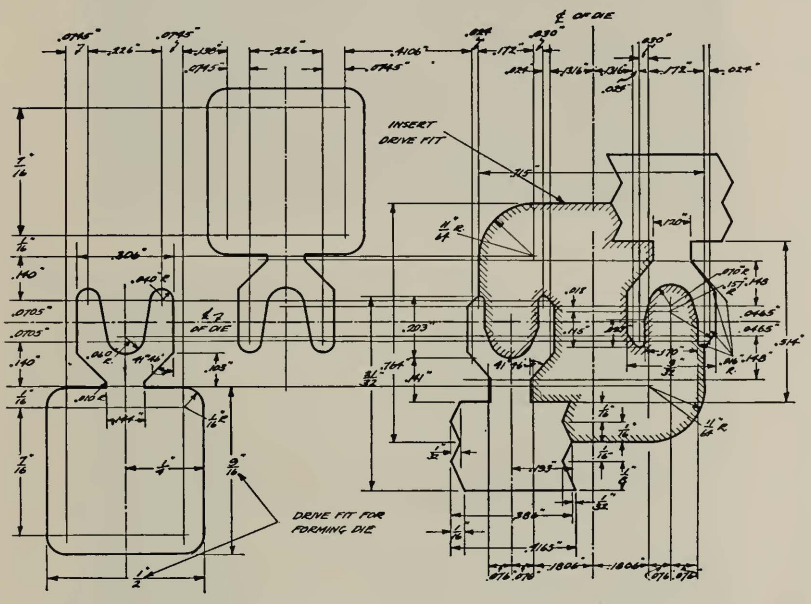


Fig. 15—Layout of die openings for solderless cord tip punch and die

the punch and die sections for the solderless cord tip punch and die previously described, will serve to illustrate the common practices followed in producing high grade work.

The first operation is the making of a tool steel master plate having all the datum or reference holes necessary for accurately locating the holes and contours of the openings in the die. The blank plate, after being squared and the sides finished parallel to each other, is mounted on the vernier milling machine shown in Fig. 14. This machine is equipped with magnifying lenses over the positioning vernier scales for adjusting the position of the table. The vernier scales read to .001 in. and by interpolation it is possible for an operator to adjust the table to within a few ten-thousandths of an inch. From the layout

of the die opening as shown on the tool drawing, Fig. 15, the required datum holes are then located by means of the vernier scales and bored to complete the master plate, as shown in Fig. 16. The



Fig. 16—Master plate for solderless cord tip die.



Fig. 17—Transferring datum holes from master plate to die block on bench lathe.

practice on master plates of this kind and other similar parts is to locate the various holes so that the error in any case is well within .001 in.

The master plate is mounted on the die block and located by means of two snug fitting pins driven into the block and projecting into the two aligning holes near each end of the plate. The die block and plate are then attached to the face plate of a bench lathe, as shown in Fig. 17. Each datum hole in turn is accurately centered with the lathe center by means of the indicator shown in the figure, the master plate removed, and the hole bored in the die block. In centering with



Fig. 18—Filing out die openings with filing attachment on bench lathe.

the indicator, the lathe is rotated and the master plate and die block shifted until the indicator pointer shows but little, if any, movement. With an indicator multiplying the movement 100 times, shifting the plate .0001 in. would move the pointer over the scale .010 in. or .0001 in. eccentricity of the hole would show a pointer movement of .020 in. The master plate gives a permanent precise outline of the important holes, radii, contours, etc., which can be utilized to considerable advantage in checking after heat treatment, and especially in making additional tools or replacement die sections.

After the boring of the holes in the die block is completed, the die openings are worked out roughly by drilling a series of holes corresponding to the shape required and brought to about .001 in. of the nominal by means of the lathe filing attachment, as shown in Fig. 18, or a standard bench filing machine. The perforating and blanking contour which is the one being filed in the illustration conforms only partially to the finished openings on the die as shown in Fig. 12, and the correct outline is obtained by means of an insert indicated on the die layout in Fig. 15. This construction is necessary because the two blanking openings are close together and could not be satisfactorily heat treated without considerable distortion. Also, it facilitates considerably the work of the tool maker in working out the die openings. The insert is heat treated before assembly in the die. The forming dies are also made separately and inserted in the square openings in the die block.

After the filing operation, the die block is heat treated and the upper and lower surfaces are then ground parallel. Although proper heat treatment is an important factor in the production of fine tools, it is too broad and extensive a subject to be considered in this paper, as the art has been developed to the point where it is now done on practically a scientific basis through the use of the most improved equipment and automatic temperature recording and control, with many different heating methods, etc., being employed for the various grades of steels and the different purposes for which they are used.

The next operation after heat treatment is the grinding of the die block openings, which is done on the bench lathe by means of the grinding attachment shown in Fig. 19. By this means the surfaces are brought to within .0003 in. to .0004 in., the most important dimensions as previously mentioned being those which affect the distance between corresponding surfaces of the blanking and forming die openings. The surfaces are then stoned or lapped by hand with about .0002 in. or less being removed as required, to give the final finish and accuracy, and the insert and forming dies, which are made with a similar degree of accuracy, fitted in place.

As can be seen from the foregoing, the highest precision work requiring the most expert workmanship comes in the final grinding, lapping, and fitting. The degree to which this must be carried, of course, depends on the requirements of the particular tool being made. In the case of the multiple bank strip and the rack tools previously described, the punch and die sections are ground to within .00005 in. of the required size and then lapped to the final dimensions, using a flat cast iron block or some other soft metal charged with an abrasive dust, such as emery, carborundum, or diamond.

An idea of what this class of workmanship means can be appreciated when it is considered that 10 degrees Fahrenheit difference in temperature will change the length of an inch block more than this .00005 in. limit. Since the change per inch in steel is about seven-millionths of an inch per degree Fahrenheit, the heat of the hands or machines may, in extremely accurate work, make sufficient difference so that

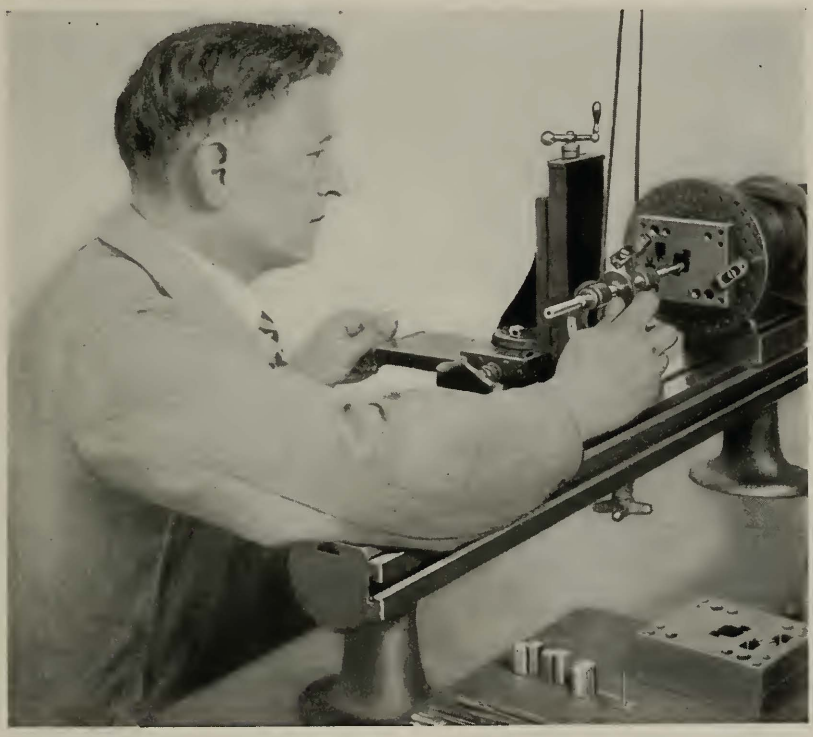


Fig. 19—Grinding die block openings after heat treatment.

parts have to be laid on steel blocks to attain room temperature before the dimensions are checked, in order that they will be of the same temperature as the master gages used. The temperature is also an important factor in producing accurate plane surfaces with a surface lap. Unless the temperature of the lap and the work is the same, a convex surface will usually be produced even though the lap itself is an accurate plane.

In making the blanking punch for the solderless cord tip tool, it is first rough milled to within 1/64 in. of the nominal dimensions,

as shown in Fig. 20. By means of a screw press, the punch is then forced into the die opening already completed, to a depth of approximately 1/64 in., and an accurate impression of the correct punch section obtained, as shown in the figure. The punch is then milled to form on the bench milling machine to within about .0005 in. to .001 in. of the nominal, the outline of the impression being used as a guide in this operation. The final shearing of the punch in the die, which amounts to practically a shaving operation, is accomplished in several steps, the excess metal being removed by filing after each operation, and the punch worked down until it enters the die to the required depth. The punch is then hardened, after which it is ground, and lapped or stoned to the exact clearance required between the punch and the die opening, which, in this case, is .0005 in. all around.



Fig. 20—Rough milled blanking punch, solderless cord tip punch and die

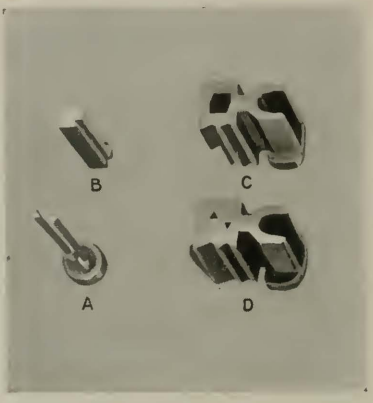


Fig. 21—Shedder, insert, and perforator punch for solderless cord tip punch and die

Fig. 21 shows "*A*" one set of the perforators for producing the two sharp projections in the cord tip stem, "*B*" the insert, and "*C*" and "*D*" the shedder before and after the insert is in place. The blanking punch also has die holes for the perforators, as can be observed from the general view of the tool in Fig. 12, and these are similarly formed by means of an insert. The perforators, which are working fits in the shedder, are .06 in. wide, $1\frac{7}{16}$ in. long and of triangular cross-section. It would, therefore, be very difficult to work out the holes straight and accurate. To facilitate the tool making work, the shedder and punch are made as shown and the insert added. This is a good example of some of the means employed for overcoming difficult tool making problems.

Making Die Blocks for Sheathing Lead-Covered Cable

The making of the die blocks used in the hydraulic presses for sheathing lead-covered cable is of interest on account of the method used. The milling of the die contour, which is irregular in shape, is done on a die sinking machine shown in Fig. 22, the upper die being

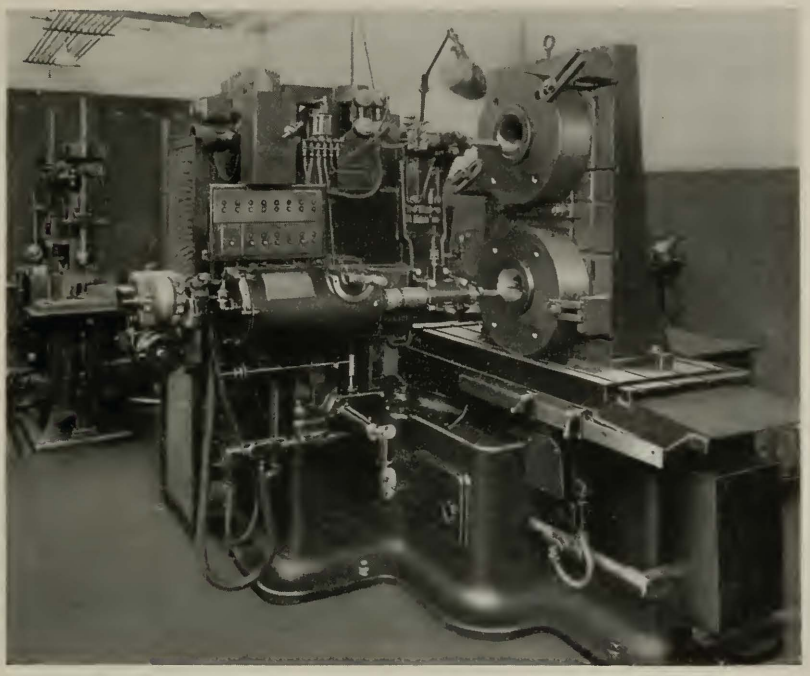


Fig. 22—Milling on die sinking machine the contour of die block for sheathing lead-covered cable

the master and the lower the die being profiled. The principle of the machine is that of having the cutter or milling tool automatically controlled and guided by means of a very sensitive tracer, which follows the contour of the master. This is accomplished by having individual motor drives and electrically operated clutches for each of the feeds, which are controlled by the movement of the tracer making and breaking the electrical circuits as it comes in contact with the surface of the master. This results in the feeds operating to move the table holding the dies and the slide holding the tracer and the cutting tool in such a manner that the tracer will follow the outline of the master die. The milling out of the opening is done by a series of either horizontal or vertical cuts, the machine automatically reversing at the

end of each cut. With this machine a set of die blocks, which would require approximately 300 hours to machine by the hand finishing method, can be completed in 80 hours. This machine was used for working out the openings of the moulding dies for the new hand set.

Precision Measuring

One of the essential factors in high grade tool work is precision measuring instruments of sufficient accuracy to check the dimensions to the limits required. The most common and practical method of making precise measurements is by comparison with standard known dimensions and most of the instruments used on a commercial basis for measuring to limits of .0001 in. or less employ this principle.

The standards used for comparison are "Hoke" or "Johanssen" gage blocks made in 81 sizes as shown in Figs. 23 and 25. The blocks are arranged in four sets, the first consisting of four blocks 1 in., 2 in., 3 in. and 4 in. in length. The second set of 19 varies from .050 in. to .950 in. in .050 in. steps. The third set of 49 varies from .101 in. to .149 in. in .001 in. steps and the fourth set of 9 varies from .1001 in. to .1009 in. in .0001 in. increments. In combination any dimension may be obtained within the limit of the set in .0001 in. steps.

The surfaces of these gages are so flat and smooth that if two or more are wrung together so as to expel the air, they will adhere to each other and resist separation at right angles to the contacting surfaces with a force of over 20 pounds per sq. in. The precision of these gage blocks at 68° F. is within .00001 in. per inch of length of the dimensions stamped on the blocks for the larger sizes and .000005 in. for the smaller sizes under one inch. Although the gages are standard at 68° F., it is of course not necessary to use them at this temperature, or make corrections when measuring metal of the same coefficient of expansion. However, as previously mentioned, it is essential that the work to be measured be at the same temperature as the gages. In addition to being used as standards for checking parts in the different measuring instruments, a variety of other uses are made of the gage block in laying out and measuring the work directly. By the use of accessories and attachments, which are furnished for holding the blocks, they may be made into inside and outside calipers, shape and height gages, etc. One of the sets which has been checked and certified by the Bureau of Standards is maintained as a standard for checking the other sets and also other master gages.

One of the most frequently used measuring instruments is the upright dial indicator gage. The part to be measured is placed on the accu-

rately lapped surface plate of the instrument and brought into contact with the vertical plunger, which operates the universal dial indicator through a lever arrangement. The movement of the pointer is about 1/16 in. for each .0001 in. vertical movement of the dial plunger. By noting the difference between the dial reading for a standard gage block of the size required and the part being measured, a comparison between the two may be made to within .0001 in. and by interpolation between the calibration marks to within a few hundred-thousandths of an inch. The universal dial indicator is also frequently used by the tool maker with a standard surface plate, and a suitable arm for holding the indicator, when checking work in process.

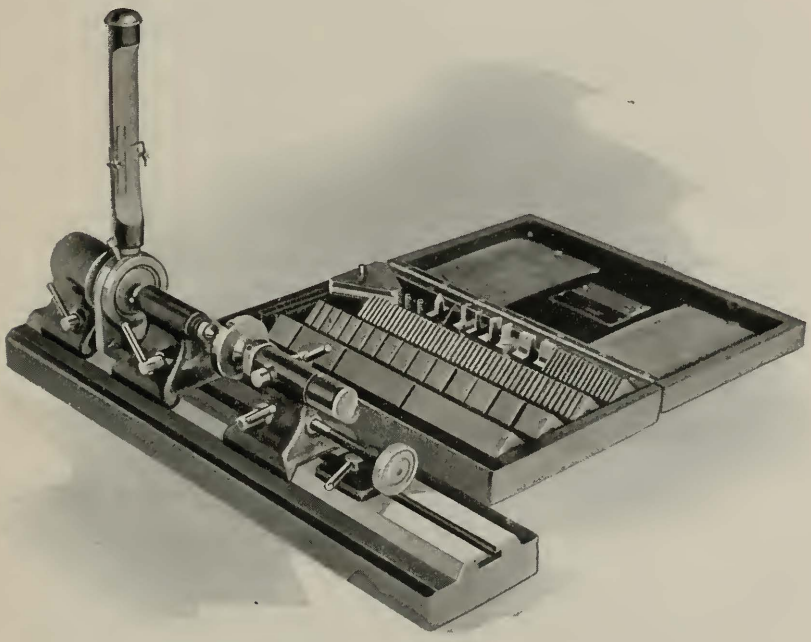


Fig. 23—Precision measuring instrument with fluid gage and "Hoke" gage block set.

If greater accuracy is required, parts are sometimes measured by comparison with the standards, using the liquid gages shown in Figs. 23 and 24. The instrument in Fig. 23, which has a multiplying ratio of 2200 to 1, was made in the Hawthorne Works Tool Room, while the other is a commercial liquid gage or prestometer. However, the comparator most generally used at present for this class of work is the optimeter shown in Fig. 25. This instrument makes use of an optical

system to magnify small measurements without the use of a vernier or other mechanical means, and due to its construction is dependable and probably less liable to variation than the liquid gage. Most of the errors due to play between mechanical parts such as gears,

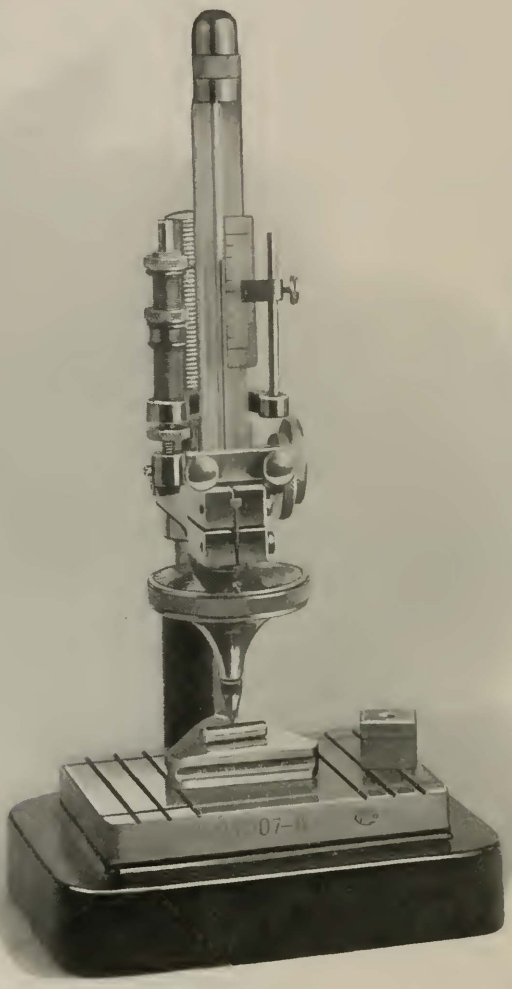


Fig. 24—Fluid gage or "prestometer" showing tool part in position for measuring

micrometer screws, knife edges, diaphragms, capillary tubes, etc., have been considerably reduced. The only moving part except the measuring feeler is a small mirror which is tilted by the upper end of the feeler and reflects the image of a stationary glass scale. The

feeler has a constant pressure of 7 or 8 ounces against the work, thus eliminating the "sense of touch" factor. Variations of the size of a part within a range of ± .0035 in. may be read directly to .00005 in. and by interpolation to within one or two hundred-thousandths of an inch.

The instrument just described illustrates the use of light for making precise measurements by the optical lever method. Another method is by the use of a lens or projection system, whereby beams of light are controlled in such a manner as to form on a screen enlarged images of objects with a high degree of geometrical similarity between the

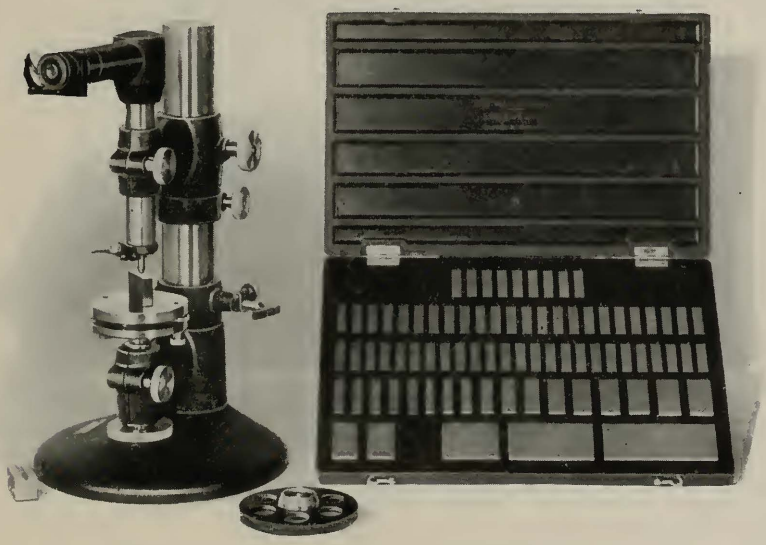


Fig. 25—Optimeter and "Johanssen" gage block set

image and the object. As the errors or variations in the object are magnified the same amount, they become correspondingly easier to observe and measure or check against a standard template, contour plate, limit chart, or accurately made scale drawing of the object. With a magnification of 250 diameters an error of only a thousandth of an inch will appear as a quarter of an inch and an error of one ten-thousandth can be readily observed. This method is particularly adaptable for measuring to close limits irregular shapes and contours, screw thread and profile gages, gear teeth, etc. The instrument used for this purpose is the contour measuring projector, the magnified image of an object being projected either on a vertical screen or the horizontal table attached to the instrument.

Two of the characteristics of light are the constancy of its wave length for a given color and its property of interference, which together establish the fact that each interference band produced by the reflection of light, as shown in Fig. 26, represents a very small definite separation



Fig. 26—Interference bands produced by the reflection of light waves

between the surfaces producing the reflections. As this separation or distance is a function of the wave length of the light used, the application of this principle permits extremely accurate measurements to within a few millionths of an inch.

Fig. 27 illustrates the application of this method in measuring a .375 in. plug, the equipment used being an optical glass flat, a metal flat on which the parts rest, a .375 in. master gage block, and a monochromatic light, usually red or green, having wave lengths of .000025 in. and .000020 in. respectively. In order to simplify calculations, the plug is placed so that its center is a distance from the gage block equal to the width of the latter. Unless the gage and the plug are exactly the same size, dark interference bands will appear across the gage block when the light falls upon it, due to the wedge-shaped air

space formed between the upper flat and the top surface of the gage block. In accordance with the theory of interference, the distance between the flat and the gage at the first band adjacent to the edge where contact between the two surfaces is made, is one half the wave length of the light used, which if red would be .0000125 in. The distance at the second band is then one wave length or .000025 in. If there are a total of three bands, the distance at the edge of the block



Fig. 27—Metal flat, optical glass flat, and gage block in position for measuring .375″ plug by light wave interference method

is .000037 in. or the difference in size between the plug and the gage is .000074 in., since the distance between them is twice the width of the block. The interference method gives a reliable and permanent unit of measurement and is probably one of the greatest refinements in precision measuring. In addition to measuring lengths, it can be used for checking the accuracy of flat surfaces, tapers, etc.

For more precise comparisons of gages and for the direct measurements of gage blocks in terms of wave lengths of light there is available a special form of the Michelson interferometer made by Zeiss, having a monochromator for selecting the particular wave length to be used, Fig. 28. This instrument is a comparatively recent development and to

our knowledge there are only two in this country at present, the other one being in the Bureau of Standards.

The light is furnished by the helium tube "*A*." The particular wave length to be used is selected by rotating a glass prism inside the case by means of the cylinder "*B*" graduated to read the wave length directly. The gage block to be measured is located at "*C*" and the interference bands are observed through the eyepiece "*E*."



Fig. 28—Laboratory interferometer

To make a direct measurement of the length of a gage, observations are made using five wave lengths, and by computing the results obtained the length may be determined to an accuracy of about two millionths of an inch. In working with this degree of precision, the instrument is used in a constant temperature room and corrections made for temperature, humidity, and barometric pressure.

Conclusion

Tool making in all its branches as carried on in the tool rooms of the Western Electric Company is a comprehensive subject, regarding which several volumes might be written if covered in detail. In the foregoing description an effort has been made to give briefly, and by considering only one branch of the work, a general picture of the high grade workmanship required and some of the equipment and instruments employed. Similar precision is required on many classes

of tools, such as jigs, fixtures, screw machine tools, milling cutters, etc., and especially gages employed in interchangeable manufacture.

This paper would not be complete without mentioning the fact that the high degree of workmanship, technique, and precision found in the product and methods of the Western Electric Company's tool rooms and many of the novel features of tool design are in a large measure due to the Works Technical Organizations operating the tool rooms. The writer is indebted to these organizations, as well as to other groups in the Manufacturing Department, for much of the material presented in this paper.

The Natural Period of Linear Conductors

By C. R. ENGLUND

SYNOPSIS: This paper describes the experimental determination of the frequency of free electrical oscillation of straight rods and circular loops. The results agree more closely with the formula of Abraham than with that of MacDonald. For three rods whose lengths were 300 cm., 250 cm. and 227.1 cm., the ratio of wave length at resonance to rod length had the values 2.11, 2.13 and 2.13, respectively. Measurements taken upon 250 cm. rods bent into circular arcs of different radii gave values of the ratio of resonant wave length to arc length which passed through a minimum value and were virtually independent of the radius of the arc over a wide range, deviating markedly only at the extreme value of minimum radius possible and infinite radius. The extreme measured range of the ratio was 2.05 to 2.166. The wave lengths were measured upon a pair of Lecher wires and a very satisfactory meter for the rapid comparison of waves of short length was found to be a quarter wave length Lecher frame. This frame showed a constant end correction so that $\lambda = 4(d + 3.1)$, d being the length of the parallel rods.

IN 1898 Abraham [1] calculated the free period of an extended but relatively narrow metallic ellipsoid of revolution when excited by an electrical impulse. To a good approximation the fundamental natural period found was related to the major axis length by the expression $\lambda/l = 2$. Obviously a rectilinear conductor of circular cross-section cannot differ markedly from such an ellipsoid and Abraham concluded that the equation $\lambda/l = 2$ was also valid for this.

In 1902 Macdonald [2] arrived, by a theoretical deduction quite different from that of Abraham, at the expression $\lambda/l = 2.53$ for the fundamental free period of a linear conductor. Moreover Macdonald assigned the same value to the linear conductor when bent into a nearly closed circle. In the next twelve years a variety of papers were published [3] giving results which were aimed at clearing up this discrepancy without however definitely settling the matter one way or the other. The subject has in recent years become of interest again following the development of the short wave vacuum tube oscillator and the conjoint use of rectilinear conductors as radiators (or "reflectors")—particularly in grids of parabolic form.

Since the universal method of measuring wave length is that of determining the nodal distances for standing waves on parallel con-

[1] Abraham, *Ann. der Phy.*, 66, 435, 1898.

[2] Macdonald, "Electric Waves," pp. 111–112.

[3] See Bibliography at end.

ductor "Lecher" systems, and since further there are practical advantages in reducing the total length of these systems to a single or half a single nodal length ($\lambda/2$ and $\lambda/4$ "Lecher" frames), the problem of the natural period of a rectilinear conductor can be broadened to include a study of the shortest favorable shape of a linear conductor for use as a wave length standard. It is the purpose of this paper to give the results of some experimental work relating to both these questions. At the same time an examination of the operation of an

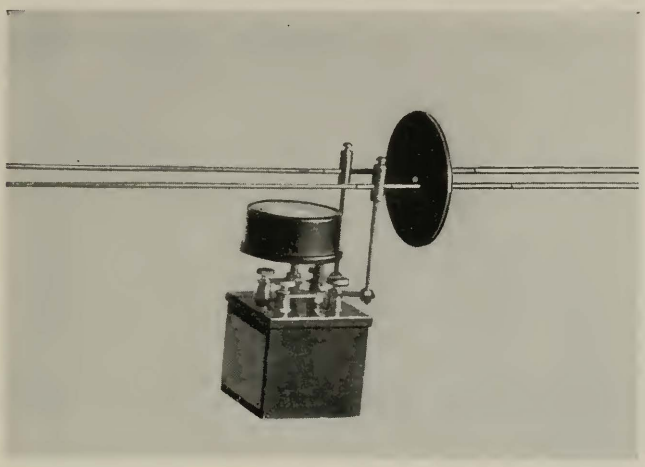


Fig. 1

extended Lecher system as a basic wave measuring apparatus was a necessary preliminary.

It was a matter of only a small amount of experimentation to demonstrate that the practical Lecher system for wave length measurement would necessarily consist of a pair of heavy uniformly spaced copper wires devoid of insulator spacers and at least a couple of wave lengths long. While the attenuation of Lecher systems made of ordinary wire is not great, as attenuations go, the accuracy with which the nodal points can be located, in the manner later described, depends markedly on the degree in which space resonance currents build up, and it is quite necessary to supply sufficient copper. The wire used here was No. 8 B. & S. gauge (3.26 mm. diam.) soft drawn copper and by "ironing" it with a slotted wood piece it was made as smooth as was necessary. It was stretched between two poles out of doors and kept tight with a turnbuckle. The spacing was fixed at 5.15 cms. by a metal bridge at one end and a micarta bridge at the other. The

total length was 15.4 meters. A photo of the sliding bridge unit is given in Fig. 1.

The method of using such a Lecher system requires consideration. If we set up a pair of parallel wires and feed energy from a generator into them, we shall have, unless the far end of the wires is terminated by the "surge" impedance of the line, a standing wave system set up. This standing wave will be most pronounced when the outer end is so terminated as to return all the energy arriving there and this requires that the terminating impedance be a pure reactance. If the extreme values of zero or infinite reactance be chosen, a current anti-node or node will respectively occur at the far end.

If the far end be reactively terminated and we observe the current distribution while moving back towards the generator, we shall find the standing wave persisting up to the generator itself. However, if the line be dissipative, the returned wave will not completely cancel the outgoing wave at phase equality locations and the current maxima and minima will become less contrasty. If the line be practically non-dissipative, the maxima and minima will not deteriorate as we approach the generator.

The returning wave is re-reflected at the generator end and traveling to the far end returns once more, this process being repeated until its amplitude has faded out. Usually the generator appears as a resistive impedance when viewed from the line so that not much energy survives reflection at this end. In any case, as the generator is the primary energy source, the generator voltage introduced into the line and the voltage of the re-reflected waves add vectorially to give a component just sufficient to maintain the standing wave line current. By an adjustment of the effective line length, either by changing the physical length, the far end reactance, or the generator impedance as viewed from the line, the power input to the line may be maximized for the particular generator used.

When this state of affairs has been attained, the standing wave may be observed by either a current or voltage operated device moved along the line. (If this device absorbs too much energy, it becomes a source of disturbing reflections itself, complicating matters by superposing on the original standing wave another pair of standing waves. It is not advisable to permit such secondary waves to exist in measurable amplitude.) Necessarily the standing wave is closely sinusoidal and at the maximum values $\partial I/\partial l = 0$ so that locating these current extremes is not an accurate experimental process. The

accuracy of determination of the distance between two consecutive values of $\partial I/\partial l = 0$ will not be sufficient unless very great care be taken, a large line current supplied, and an indicator responding well to $\partial I/\partial l$ (such as a square law thermocouple) be used. For the attainment of greater accuracy an average over a number of nodal distances must be used. In short, this method of measurement requires a relatively long line for accuracy, up to the limit where a deterioration of the maxima and minima has become pronounced due to attenuation. At the current minima $\partial I/\partial l$ is great enough for good settings but no meters of requisite sensitivity at the zero end of their scales exist.

Another and more sensitive method of observation of nodal distances is to make use of the variation of line current as the total line length is varied, particularly if both ends of the line be pure reactances and the line conductors have an adequate copper content and very good insulation. Coupling such a line weakly to a generator by merely placing the generator in the neighborhood makes it possible to build up very sharply resonant standing waves so that settings without any particular precautions can be made to one part in 3,500. Of course the point located is again one where $\partial I/\partial l = 0$ but the value of ΔI, for a given value of Δl, is very much larger. Moreover the accuracy is not decreased by shortening the line [4] and, since the resonance energy is then dissipated in a shorter length of line, the corresponding increase in the resonance current makes the antinode easier to locate. Although this method is very sensitive to energy losses it is by far the best method of using a Lecher system. Essentially it is nothing but a sharp "tune" observed and interpreted as space resonance. In the present work the length of the Lecher system was sufficient to observe four resonance maxima throughout the range of wave lengths used, giving, by difference, three wave length readings. These readings could readily be duplicated to one part in 3,500; it is improbable however that the velocity of propagation along the wires is within one part in 3,500 of that of free space so that this precision is unusable, though comforting. This accuracy of setting would be useful where small frequency differences or line constant changes are to be observed. Since the measurements checked admirably from day to day and the line attenuation was low, it was assumed that the line velocity was not affected by the adjacent ground (110 cms. at lowest point) and was near enough to that of the velocity of light to allow the line to be used as the basic wave length standard.

[4] A. Hund, Sci. Paper No. 491, Bureau of Standards, 1924, points out the same fact.

Resonance Period of a Straight Rod

It was at first thought that it would be relatively simple to set up a vacuum tube generator together with a rectilinear rod and run resonance curves on the latter. This did not prove to be the case however. Working indoors was impossible and all the apparatus had to be moved out of doors. When the two antennas (the generator

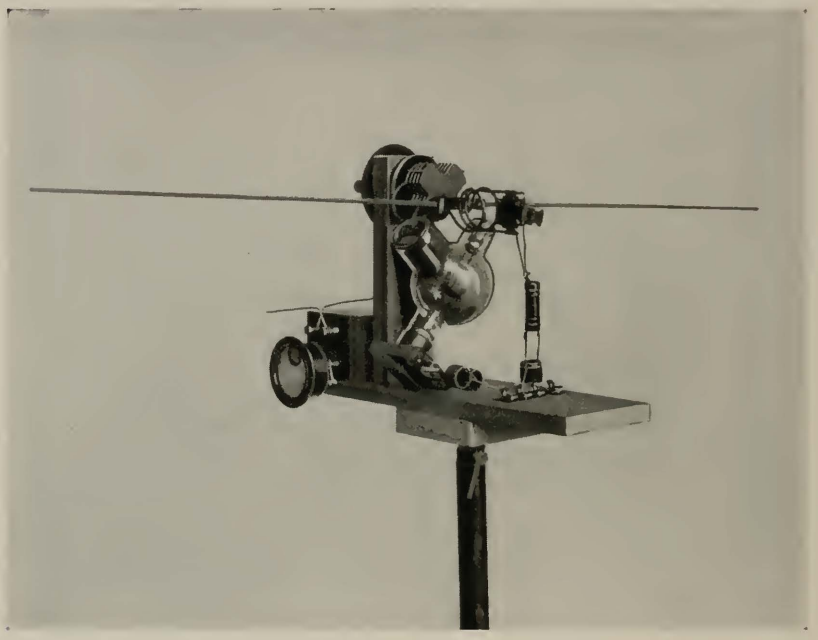


Fig. 2

antenna and resonant rod) were mounted vertically, the operator became a mobile reflector himself seriously disturbing the transmission. Moreover the ground was unsymmetrically disposed with respect to the two antenna ends and this was felt to be a disadvantage. With the antennas horizontal these objections vanished but the reflection from the ground had to be taken into account. This latter was the arrangement finally adopted and is shown in Fig. 3, the apparatus in question being mounted on the two tripods.

The attempt was made to get a generator whose radiation field was constant over a wide wave length range so that the rod resonance wave length could be obtained by observation while turning the generator tuning condenser. As a matter of fact, the generator output was apparently satisfactorily constant while actually not so,

and some time was wasted trying to get consistent results. Finally a control meter was placed on the generator and resonance curves run, observing by small steps wave length, rod meter deflection, and control meter deflection. Then by reducing the rod meter deflection to a standard control meter value satisfactory observations were obtained. As it turned out, the averaged value of all the unsatisfactory observations checked the resonance curve value very closely, but the individual observations scattered all over the rather broad resonance curve top. To avoid the reaction of antennas upon each other they must be separated by at least a wave length, and such a

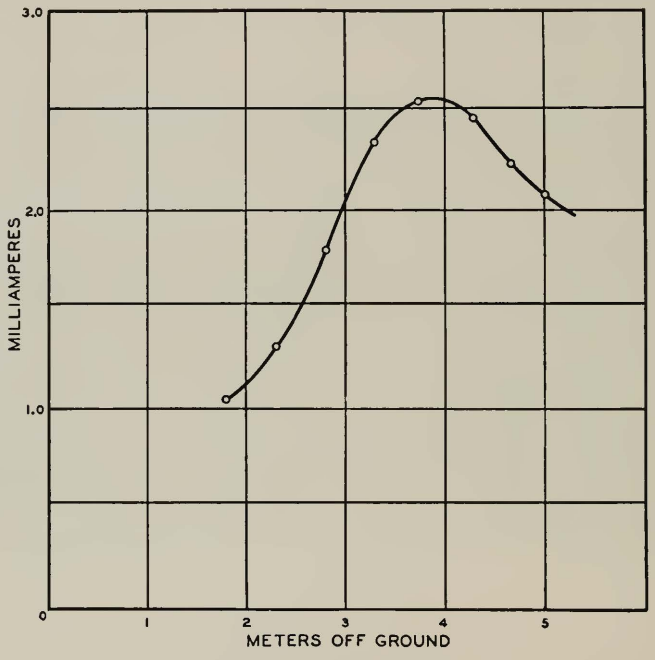


Fig. 3

separation here resulted in too low field strengths unless the antennas were off the ground sufficiently to get an additive combination of the direct and earth reflected radiations. The resonant rod should in any case be well off ground to make certain that its period is not affected by the ground. Check tests showed that at 4 meters distance the ground did not affect the free period markedly. It would be much simpler to observe the resonance curve of a variable length rod, at constant wave length, but it would not be permissible to use a rod of variable diameter and a telescoping arrangement is the only practical method of obtaining collapsibility.

The generator used was an UX852 tube connected as in Fig. 2. By means of the variable condenser shown a wave length range of 4.24 to 8.44 meters was obtained. This condenser is a cut down "Remmler," the oscillating coil is a three turn center-tapped unit

of 1/8 inch (0.32 cm.) copper tubing, the choke coils are 10 microhenry units resonant at 6.1 meters and the antenna rods are connected directly across the resonant circuit. This apparatus was finally set up on a tripod raising the antenna 2.55 meters above ground. The control meter consisted of a pair of 15 cm. wires connected to a thermocouple and Weston model 301 micro-ammeter combination. It was not resonant in the generator range and was fastened on the generator base permanently.



Curve I—300 cm. rod

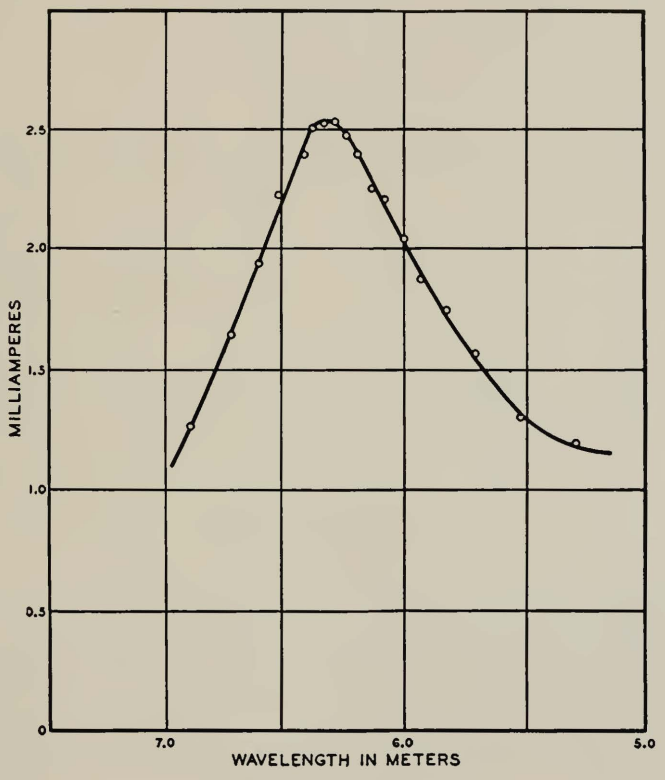
The rods studied were ordinary 1/2 inch copper and brass tubes and were straightened and mounted in a bracket consisting of a pair of phosphor bronze knife edges spaced 27.5 cms. apart. These knife edges were attached to a micarta strip and connected to a 600 ohm thermocouple and Weston model 301 micro-ammeter combination as for the control meter. A tripod supported the rod mount and the meter was read with a field glass at the generator site. The wave length was determined by a $\lambda/4$ Lecher "frame" (see Fig. 5) calibrated by means of the Lecher wires. This will be described later.

A full scale deflection of the resonant rod micro-ammeter corresponded to 4 milliamperes heater current and this was shown, by

means of experiment with two rods each equipped with a removable thermocouple, to have no effect on the resonant frequency. With two rods in proximity the resonance was much more sharp and definite than for one rod, and was also more sensitive to effects producing variations in the natural period.

Three separate rods were used, their specifications being,

Length	O.D.	I.D.	Material
300 cm.....	1.27 cm.	1.02 cm.	Copper
250 "	1.27	1.02	"
227.1 "	1.27	1.02	Brass



Curve II—300 cm. rod

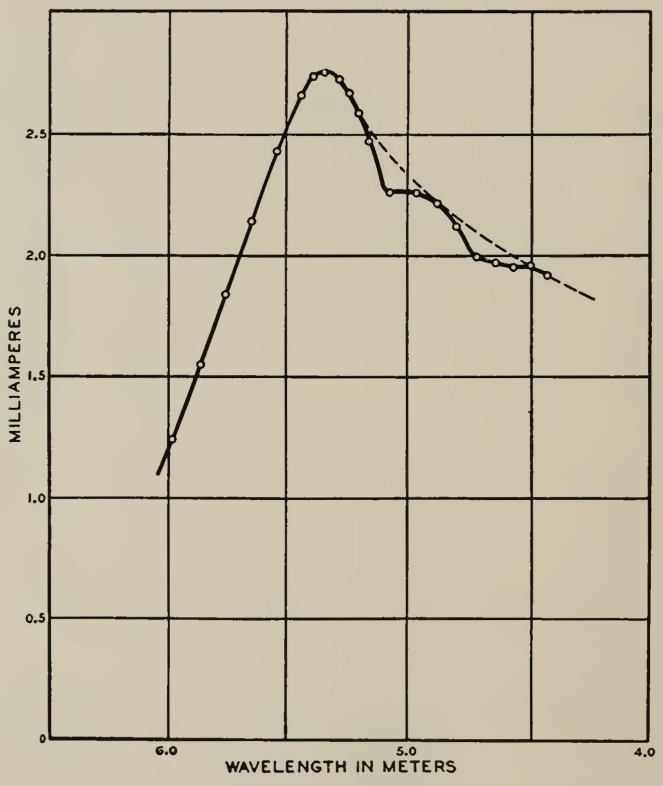
Curve I shows the maximum current versus height above ground relation for the 300 cm. rod. The generator antenna was 2.55 meters above ground and the horizontal spacing between generator antenna and resonant rod always 7.7 meters, both antennas horizontal. The

curve thus needs correction for the fact that the air line distance varied as the rod was lowered and raised. This correction is small however.

It will be observed that the maximum current occurs at 3.85 meters and taking this value for determining the distance to the reflecting ground surface gives, if we assume a radiation field only,

$$(1.3 + 2h)^2 + (770)^2 = \left(\sqrt{130^2 + 770^2} + \frac{\lambda}{2}\right)^2$$

or with $\lambda = 6.34$ meters, $h = 3.26$ meters. The difference $3.26 - 2.55 = 0.71$ meter is the apparent distance under ground of a metallic sheet equivalent to the ground itself.



Curve III—250 cm. rod

Curves II and III give the resonance curves of the 300 and 250 cm. rods respectively, mounted horizontally and 4 meters above ground. The latter curve is complicated by two extraneous resonances

occurring somewhere in the generator circuit. With the advent of inclement weather the source of these resonances could not be looked for. These dips in the resonance curve are not evident to an observer watching the resonance rod meter as the generator wave length is varied; they become evident only on plotting a carefully taken resonance curve. The results both for preliminary eye settings and final resonance curve are:

Rod	By Res. Curve		By Eye Settings	
	λ	λ/l	λ	λ/l
300	6.34	2.11	6.33	2.11 (av. 10 settings)
250	5.36	2.14	5.32	2.13 (av. 13 ")
227.1	—	—	4.84	2.13 (av. 8 ")

A check eye setting on the 250 cm. rod mounted vertically agreed with the other eye settings. The eye settings, or eye estimates of the top of the rod resonance curve, were made first, the resonance curves were run last. On discovering the "dips" in the 250 cm. rod curve it became useless to run a 227.1 resonance curve before eliminating these dips, the preliminary curve being discarded. It is not certain however that these "dips" were present in most of the eye settings as these were chiefly made with the generator nearer ground and with shorter power leads. They are given for what they are worth.

Evidently experiment more nearly checks Abraham [5] than Macdonald, the rods operating as if their effective lengths were 6–7 per cent greater than their physical lengths. Whether this "end correction" varies with the rod diameter was not investigated owing to bad weather. Several rods were however bent into circles, cut to 250 cms. perimeter (outside length), and their natural periods determined. The results follow below. With the bending, the radiation resistance of the rods went down pronouncedly and the knife blade contacts were shortened to span a distance of 10.2 cms. A preliminary test showed the natural period of such rings to be independent of their orientations; they were therefore hung up in the knife edges with open gap downwards. The top edge of the ring was always

[5] Abraham gives a correction term of the form

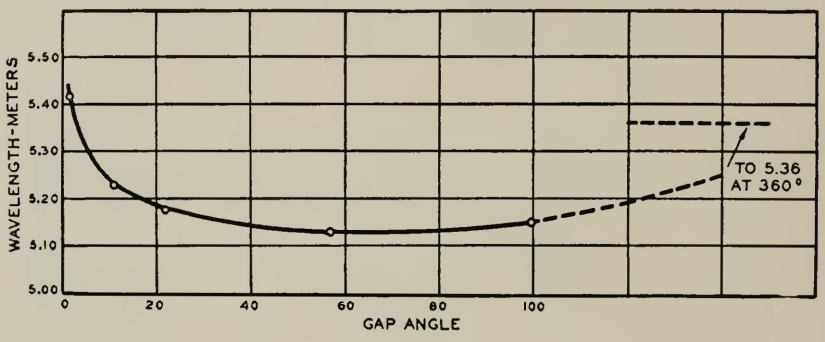
$$\frac{\lambda}{l} = \frac{2}{n}\left[1 + C_n \cdot \left(\frac{1}{4 \log \frac{2l}{d}}\right)^2\right],$$

where "n" is the order of the harmonic and C_n a complicated integral. For "n" = 1 and the 300 cm. rod, $\lambda/l = 2.018$, which is too small.

3.17 meters above ground, generator antenna horizontal and 2.55 meters above ground, horizontal separation between ring plane and transmitting antenna 7.2 meters. The results were:

RING SHAPED 250 CM. ROD WITH GAP

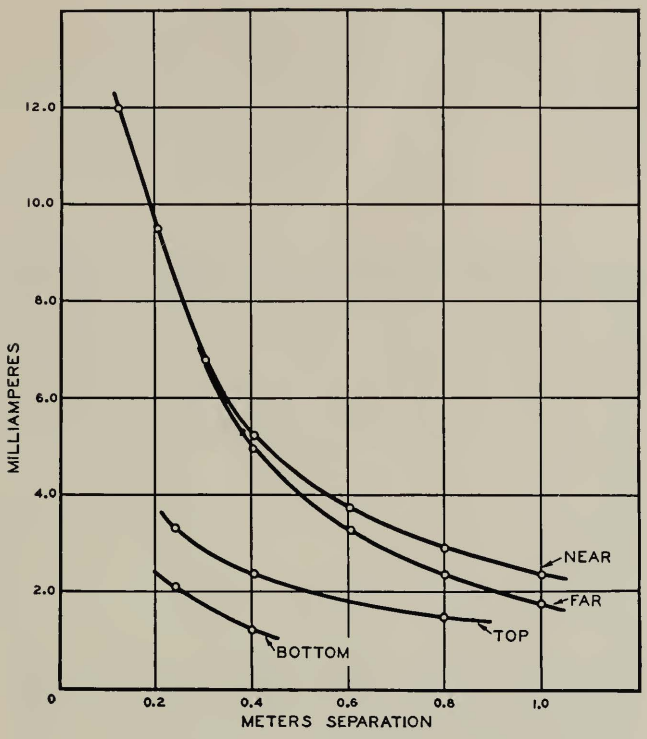
Gap Angle	Ring Radius	λ Meters	λ/l
1.58°	40. cms.	5.414	2.166
10.8	41.1	5.228	2.09
21.8	42.4	5.178	2.07
56.4	47.2	5.128	2.05
99.8	55.1	5.148	2.06
360.0	∞	5.36	2.14



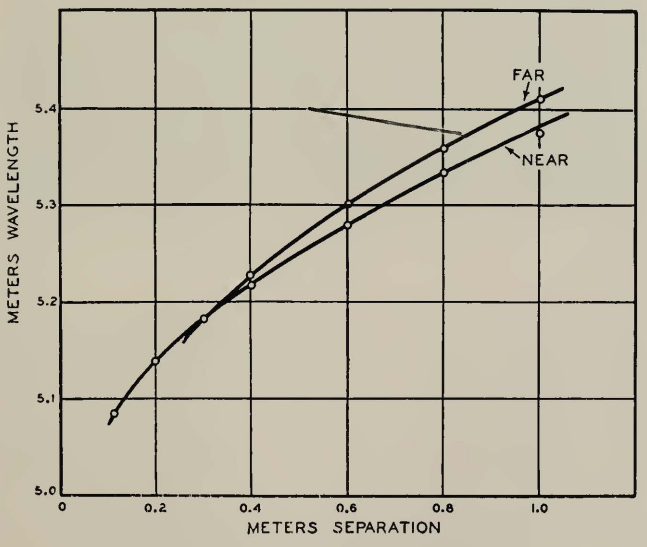
Curve IV

These values are plotted in Curve IV and do not check Macdonald's conclusion earlier referred to. In fact the first part of the curve is most simply explained by viewing the rod as a resonant inductance whose tuning capacity is decreased as the rod gap opens. The existence of a minimum resonant wave length is less easy to explain in this manner.

It was early observed that the current amplitude and sharpness of resonance of a rectilinear rod were greatly increased by the proximity of a second rod. Curves were therefore run with two parallel rods arranged both in horizontal and vertical planes, the spacing being changed while current magnitude and resonance wave length were observed. In each case the rods were symmetrically mounted about a central point held at the height of the generator antenna above ground (2.55 meters) and the short knife edge support, mentioned in connection with the rings, was used. All the antennas were horizontal and as accurately parallel as it was possible to set them. Fig. 4 shows one mounting and Curve V the current versus spacing relation, while Curve VI gives the resonance wave length versus spacing.



Curve V—Two rods each 250 cm. long



Curve VI—Two rods each 250 cm. long

It is unwise to attempt any critical conclusions from these curves as long as the standing wave pattern of the direct and earth reflected wave interference is not known. But two facts seem clear, viz.: a closely spaced rod pair gives a marked current step up over that of a single rod, and the natural period of a rod pair approaches the value $\lambda = l/2$ as the spacing decreases. It is obvious that the currents



Fig. 4

in the two rods, at close spacing, are nearly anti-phased and that their vector sum must be nearly that current which an isolated rod would carry. The analogy with an anti-resonant circuit is evident, the "stepped up" current being limited only by the ohmic losses in the rods. Actually the tune, at close spacings, was excessively sharp and hard to set for with the generator condenser. The exigencies of the mounting of the rods and the observing of the meter prevented observations at close spacing for the "far," "top" and "bottom" meter positions.

High Frequency Wave Meter

A pair of Lecher wires constitutes a wave measuring system much too awkward and extended for rapid use, and some apparatus much more portable and speedy in operation is necessary; especially when running resonance curves. Such an apparatus is a pair of heavy uniform and parallel conductors arranged with one or two sliding

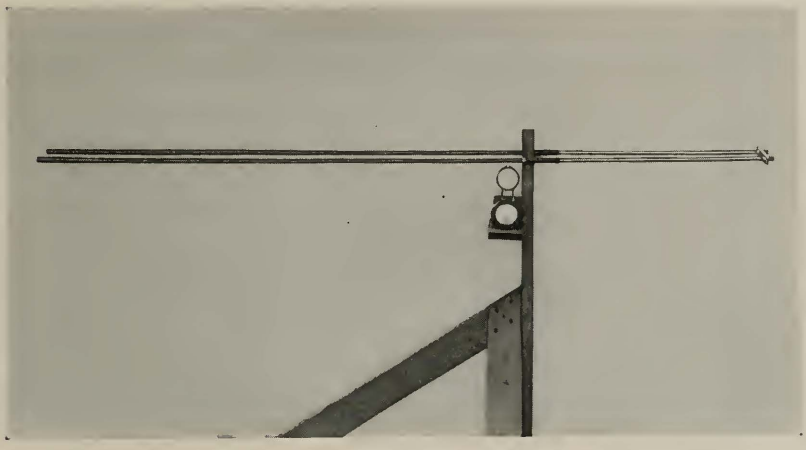


Fig. 5

metallic short-circuiting discs so as to constitute a quarter or a half wave length "Lecher frame" (see Fig. 5). Such a frame, if containing sufficient copper, is very sharply resonant, need only be a quarter wave length long, and after calibration becomes a wave length meter indicating to a precision of 1 part in 2,500 with the greatest ease. As an accessory apparatus such a wave frame was constructed, calibrated, tested for factors affecting its accuracy and used for most of the measurements reported above.

The Lecher frame shown in Fig. 5 was made of a pair of straight copper tubes 1.27 cms. diameter spaced 10.1 cms. center to center and sliding through a brass disc 15.5 cms. diameter, 0.3 cm. thick, with inserted guide tubes. It had a workable wave length range of 4 to 7.5 meters and the resonance setting was indicated by a three turn coil-thermocouple-microammeter combination placed with coil clearing one of the rods at the disc end by approximately two cms. It was ordinarily cleaned to make good contact at the disc guides but at no time was any indication noticed of an apparent lengthening of the frame due to a moving back inside of the guides of the contact point. It was calibrated over its whole range in terms of the Lecher

wire system already mentioned, calibrated not once but various times and unfailingly indicated 3.1 cms. too short. That is, the distance "d" from open end to disc was 3.1 cms. short of $\lambda/4$, or $\lambda = 4(d + 3.1)$. It was not found possible to make a trombone slide which maintained its spacing accurately, and before each reading was completed a paper template was laid on the open end and the tubes given a slight bend to obtain parallelism. The setting was then completed. This process was much less bothersome than might appear from its description.

A 35 x 44 cm. copper plate was clamped to the brass disc to increase its effective area. This brought the 3.1 cm. correction down to 2.67 cms. (measured at 5.29 meters wave length). The end effect is therefore chiefly due to the open end. No doubt it will change if the spacing of the Lecher frame is changed; the fortunate feature is that it appears constant for a given spacing.

To obtain an idea of the effect of a slight degree of non-parallelism and the presence of insulation material, the rods were first bent out, then in, next a 1.3 x 3.8 x 12.7 cm. hard rubber block was laid flat across the outer end and finally this block was laid across the middle of the frame. The results were:

Lecher Frame Length	Condition	True $\lambda/4$ (Av. of Four Settings Each)	Deficiency in Frame
122 cms.	Parallel	125.07 cms.	3.07 cms.
"	Diverge 0.2 cm.	125.04	3.04
"	Converge 0.2 cm.	125.20	3.20
"	Rubber plate at end	126.21	4.21
"	Rubber plate at middle	125.58	3.58

Obviously a slight divergence is an advantage rather than otherwise, for an uncalibrated frame, and dielectric spacers at high potential points are sources of noticeable error.

BIBLIOGRAPHY

Abraham, M. Ann., 2, 32, 1900. Elektrische Schwingungen in einem Frei Endingenden Draht.

Kiebitz, F. Ann., 5, 872, 1901. Über die Elektrischen Schwingungen eines Stabförmigen Leiters.

Willard and Woodman. Physical Rev., 18, 1, 1904. A study of the radiations emitted by a Righi vibrator.

Rayleigh. Phil. Mag., 8, 105, 1904. On the electrical vibrations associated within terminated conducting rods.

Pollock. Phil. Mag., 7, 635, 1904. A comparison of the periods of the electrical vibrations associated with simple circuits.

Cole, A. D. Phys. Rev., 20, 268, 1905. The tuning of Thermoelectric receivers for electric waves.

Blake and Fountain. Phys. Rev., 23, 257, 1906. Reflection of electric waves by screens of Resonators and by Grids.

IVES, J. E. Phys. Rev., 30, 199, 1910. The wave length and overtones of a linear electrical oscillator.

BLAKE AND RUPPERSBERG. Phys. Rev., 32, 449, 1911. On the free vibrations of a Lecher system using a Blondlot oscillator.

BLAKE AND SHEARD. Phys. Rev., 32, 533, 1911; 35, 1, 1912. On the free vibrations of a Lecher system using a Lecher oscillator.

SEVERINGHAUS AND NELMS. Phys. Rev., 1, 411, 1913. Multiple reflection of short waves from screens of metallic resonators.

NELMS AND SEVERINGHAUS. Phys. Rev., 1, 429, 1913. Resonators for short electrical waves.

ARKADIEW, W. Ann., 45, 133, 1915. Über die Reflexion Elektromagnetischen Wellen an Drahten.

ABRAHAM. Jahr. d. D. T. u. T., 14, 146, 1919. Die Strahlung von Antennen Systemen.

DUNMORE AND ENGEL. Bur. of Stand. Sci. Paper No. 469, Apr. 11, 1923. Directional radio transmission on a wave-length of ten meters.

TATARINOFF. Jahr. d. D. T. u. T., 30 (Oct.), 1927.

The Measurement of Capacitance in Terms of Resistance and Frequency

By J. G. FERGUSON and B. W. BARTLETT

SYNOPSIS: The adaptation of a bridge circuit due to M. Wien together with apparatus and procedure is described which permits measurement of capacitance in terms of resistance and frequency with an accuracy comparable to that of the primary standards. Among its advantages over the Maxwell method commonly employed are the use of a single frequency voltage and the fact that there is no general limitation placed on the type of condenser which may be measured or on the frequency at which the measurement may be made. The method is also applicable to the determination of inductance since its unit, like that of capacitance, may be derived from the units of resistance and frequency.

INTRODUCTION

CONDENSERS are commonly measured by comparison with standard condensers of known value by means of one or another of the well-known bridge methods. The accuracy with which such measurements can be made depends upon the accuracy with which the capacitance of the standard is known.

The unit of capacitance is derivable from those of resistance and frequency and to obtain an absolute value for a standard of capacitance, some method is required for a precise determination of capacitance in terms of frequency and resistance. Of the methods which have been proposed, few yield the accuracy with which the primary standards of resistance and frequency are known and reproducible.

A generally accepted method for the absolute determination of capacitance in terms of resistance and frequency is to use a bridge, due to Maxwell,[1] employing the alternate charge and discharge of a condenser. This method has been used successfully by the Bureau of Standards,[2] which has obtained results of high accuracy. Several fundamental limitations, however, make it difficult for general use. Because of the operation of charge and discharge it is only applicable to the measurement of capacitances which are independent of frequency.[3] Practically this limits the method to the measurement of air condensers, which in large sizes are not very stable. Moreover the balance depends on the integration of successive charges and discharges of a condenser through a galvanometer and great care is required to insure that the galvanometer integrates correctly.

[1] J. Clark Maxwell, "Electricity and Magnetism," second edition, Volume 2, pp. 776–7.
[2] E. B. Rosa and N. E. Dorsey, *Bureau of Standards Bulletin*, Vol. 1, p. 153.
[3] H. L. Curtis, *Bureau of Standards Bulletin*, Vol. 6, 1910, p. 433.

The present paper describes the adaptation of a bridge circuit due to M. Wien,[4] together with apparatus and procedure, which permits a measurement of capacitance in terms of resistance and frequency with an accuracy comparable to that of the primary standards. To illustrate the possibilities of the method in practice the results of a specific determination are included. Among its advantages over Maxwell's method are the use of a single frequency voltage and the fact that there is no general limitation placed on the type of condenser which may be measured or on the frequency at which the measurement may be made.

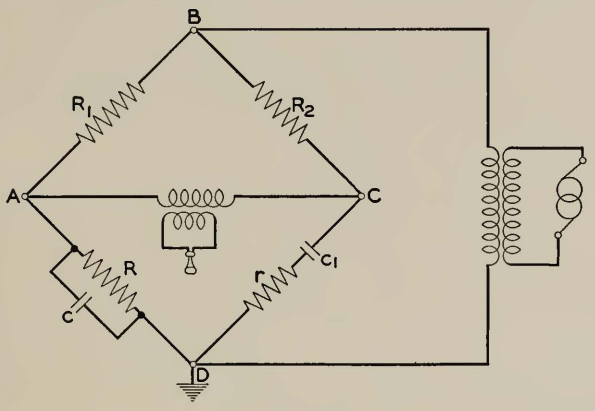


FIG. 1

The method described is also generally applicable to the determination of inductance, since its unit, like that of capacitance, may be derived from the units of resistance and frequency. The circuit and procedure to be described may be used with a change of only minor details.

In its simplest form the bridge, as shown in Fig. 1, consists of two equal resistance ratio arms, a third arm containing a capacitance and a resistance in series, and a fourth arm containing a capacitance and a resistance in parallel. A balance is easily made by varying any two of the five variables, viz., the two capacitances, the two resistances associated with them, and the frequency.

If, at balance, the frequency and any two of the other variables are known, the remaining two can be determined. Thus if the frequency, the resistance in the series arm, and the resistance in the parallel arm are known, the magnitude of both capacitances can be determined. However, the equations for balance, which are given below, are such that if the ratio of the capacitances and the value of the frequency are

[4] M. Wien, *Weid. Ann.*, 1891, p. 689.

known, the magnitudes of the capacitances can be determined from the knowledge of one only of the resistances, e.g., that in the series arm. Since the ratio of any two capacitances may be obtained with a high degree of precision by supplementary measurements, it therefore becomes possible to use the bridge just described without a knowledge of the parallel resistance, the measurement of which presents certain practical difficulties.

Although the Wien circuit is fundamentally simple, it is subject to many severe requirements when used to make an accurate determination of capacitance in terms of resistance and frequency, and must embody in its construction the refinements necessary for work of such high precision. In the Bell Telephone Laboratories there is available a capacitance bridge [5] of high precision, which is ordinarily used for the direct comparison of capacitances, and which, with slight modifications, is readily adapted to this purpose.

Theory of the Circuit

If in Fig. 1 the ratio arms are equal in resistance and phase angle, the equation of balance may be written

$$\left(\frac{1}{R}+j\omega C\right)\left(r+\frac{1}{j\omega C_1}\right)=1,$$

and separating reals from imaginaries

$$\frac{C}{C_1}=1-\frac{r}{R} \tag{1}$$

and

$$CC_1=\frac{1}{rR\omega^2}\cdot \tag{2}$$

From these two equations it is obvious that, if the values of r, R and ω, are known, the true values of C and C_1 can be determined.

However, the method of calibrating the capacitance bridge, which is described below and which is carried out irrespective of this determination gives the value of $\frac{C}{C_1}$, precisely, and this allows the reduction of the quantities to be determined to two, ω and R or ω and r.

Let C and C_1 now be taken as the values of the two condensers as measured on the capacitance bridge to determine their ratios $\frac{C}{C_1}$.

[5] G. A. Campbell, *Elect. World and Engineer*, April 2, 1904; *Bell System Technical Journal*, July 1922; W. J. Shackelton and J. G. Ferguson, *Bell System Technical Journal*, Jan. 1928.

Since this ratio as measured is the true ratio, both of the measured values must be multiplied by the same factor to give the true values, and the following substitutions may be made in formulæ (1) and (2):

$$KC_1 \text{ for } C_1$$

and

$$KC \text{ for } C,$$

where K is the correction factor necessary to reduce the values measured on the bridge to their true values. The formulæ now become

$$\frac{C}{C_1} = 1 - \frac{r}{R}$$

and

$$K^2CC_1 = \frac{1}{rR\omega^2},$$

from which, eliminating R by the use of the ratio $\frac{C}{C_1}$

$$K = \frac{1}{rC_1\omega}\sqrt{\frac{C_1 - C}{C}}.$$

In the foregoing C and C_1 are assumed to be pure capacitances, and r and R pure resistances. Of course in practice neither pure capacitances nor pure resistances are obtainable. The former will have some slight conductance and the latter some slight reactance. It we use condensers having small losses, and resistances having small phase angles, the conductance of the condenser C (Fig. 1) may be considered as a resistance in parallel with R, and that of C_1 as a resistance, r_1, in series with r. Similarly, the reactance of R may be considered as a capacitance C', either positive or negative, in parallel with C, and the reactance of r as a capacitance, C_1', in series with C_1. Of these quantities the conductance of C may be neglected, since the use of ω, r, and the ratio $\frac{C}{C_1}$ as parameters eliminates R from the formula for K, and hence it is unnecessary to know it exactly. Including these second order quantities the formula for K becomes, using the notation above,

$$K = \frac{1}{(r + r_1)\left(\frac{C_1C_1'}{C_1 + C_1'}\right)\omega}\sqrt{\frac{\frac{C_1C_1'}{C_1 + C_1'} - (C + C')}{C + C'}}.$$

Now in the range of impedances actually used in the following determinations of K it was readily possible to obtain resistance units for r in

which the reactance was so small that $\frac{C_1C_1'}{C_1 + C_1'}$ was no different from C_1 to the order of accuracy of the determinations. The parallel capacitance of R in the cases where single unit resistances only were used could also be made negligibly small compared with C in some cases, though the resistance values of R were in general considerably higher than those of r, and it was therefore more difficult to secure very small phase angles in the former. However, in a large number of the determinations a shielded resistance box was used for R, its phase angle was some 5 to 10 times that of the single units, and too large to neglect. Accordingly C_1' can be eliminated from the formula for K for the purpose of this investigation while C' cannot. The formula may then be written in the more simple form:

$$K = \frac{1}{(r + r_1)C_1\omega}\sqrt{\frac{C_1 - (C + C')}{C + C'}}. \tag{3}$$

The nominal values of the first order quantities used in the actual determinations are shown in Table 1. In this table Q is the ratio of reactance to resistance of either of the total arm impedances r and C_1 or R and C.

TABLE I

NOMINAL CAPACITANCE AND RESISTANCE COMBINATIONS USED IN DETERMINATION OF K

C_1 $\mu f.$	r ohms	C $\mu f.$	R ohms	f cycles	Q
.4	690	.1	920	1000	.6
.2	800	.1	1600	1000	1.0
.4	400	.2	800	1000	1.0
.1	1000	.072	3520	1000	1.6
.2	1000	.078	1640	1000	.8
.3	1000	.066	1280	1000	.6
.4	1000	.055	1160	1000	.4
.1	1000	.039	1630	2000	.8
.2	1000	.027	1160	2000	.4
.3	1000	.020	1070	2000	.3
.4	1000	.015	1039	2000	.2
.1	500	.091	5500	1000	3.2
.2	500	.143	1750	1000	1.6
.3	500	.158	1055	1000	1.2
.4	500	.154	810	1000	.8
.1	500	.072	1760	2000	1.6
.2	500	.078	820	2000	.8
.3	500	.066	640	2000	.6
.4	500	.055	580	2000	.4

$$K = \frac{1}{\omega C_1 r}\sqrt{\frac{C_1 - C}{C}}.$$

As mentioned above the method of this paper may obviously be extended to the measurement of inductance in terms of resistance and frequency. If in the circuit of Fig. 1, C_1 is replaced by an inductance L_1 and C by an inductance L, at balance

$$L = \frac{r^2 + \omega^2 L_1^2}{L_1 \omega^2}. \tag{4}$$

If as before L_1 is known in terms of L, i.e., if

$$L_1 = AL,$$

L_1 can be eliminated from (4) and

$$L = \frac{r}{\omega}\sqrt{\frac{1}{A(1-A)}}. \tag{5}$$

Expressing the relation (5) in terms of K, as in (3) above,

$$K = \frac{r}{\omega}\sqrt{\frac{1}{L_1(L-L_1)}}. \tag{6}$$

This formula neglects the effective resistance of L_1. The accuracy with which the value of K can be determined will in practice probably depend upon the accuracy with which the effective resistance of the inductance L_1 can be determined, which in general will be somewhat less than the accuracy of determining the conductance of the corresponding capacitance.

Description of Apparatus

The bridge equipment was a completely shielded equal-ratio bridge built for the comparison of capacitance and including standard condensers in the bridge itself. In its adaptation to the Wien circuit the standard condensers were cut out leaving a pair of equal-ratio resistance arms, properly shielded. The two additional arms were made up of external resistances and the condensers being measured.

A description of the arrangement of the standards in the capacitance bridge, however, is necessary to explain the means of obtaining the precise value of the ratio of any two capacitances. The capacitance standards, self-contained in the bridge, are variable from 0 to 1 μf and are arranged in decade form, first an air condenser with a range of slightly more than 10$\mu\mu f$, then fixed condensers up to 1 μf in 5 additional decades, each consisting of unit condensers controlled by 10 point switches. An external capacitance is measured by turning the

dials of the standard capacitance until a balance is obtained and the value is then read from the dial settings and the reading of the air condenser, which has a minimum scale division of .2 $\mu\mu f$.

The bridge condensers cannot be made exactly direct reading, and for accurate work the bridge must be calibrated. This calibration may be made very simply due to the fact that the maximum setting on any dial is approximately equal to one step on the next higher dial. By the use of an auxiliary external condenser it is possible to get a balance with any desired setting of the bridge. Thus the maximum of one dial may be compared with each individual step of the next higher dial by balancing the bridge first with the maximum setting of the lower dial and then with that dial set at zero and the next higher dial moved up one step, no change being made in the auxiliary condenser. The change in capacitance required for balance, that is, the difference between the dial settings, is read on the air condenser. Since the condensers in each decade are completely shielded from those in the other decades this procedure gives an accurate comparison of the ten steps of any dial with one another, and with the maximum setting of the next lower dial. Evidently an extension of this method will furnish a precise comparison of any bridge setting with any other, although it gives no information as to the absolute values of any of the settings.

In practice after the above "step-up" calibration, as it is called, is performed the values of all the bridge condensers are computed in terms of an assumed value of a single one. This furnishes a bridge calibration of which the consistency is dependent only on the accuracy of the "step-up" and of which the accuracy is dependent only on the value of the single calibrating standard. In general the assumed value of the calibrating standard will be in error, its true value being a constant, K, times its assumed value. Any reading on the bridge using the calibration will, therefore, require a correction by this same factor K. Now let us suppose that this bridge has exactly equal ratio arms, is calibrated as described above and is used to measure successively two capactitances whose measured values are found to be C and C_1. Their true values will then be KC and KC_1 and their ratio will be $\frac{C}{C_1}$, which is the true ratio between the capacitances irrespective of the value of K, that is, irrespective of the absolute accuracy of the measured values. By means of this type of precision capacitance bridge the ratio between any two capacitances may thus be obtained regardless of the absolute accuracy with which the capacitance of either is known.

Actually the best known value is always assumed for the capacitance used as the standard in computing the calibration of the bridge, and

accordingly the constant K by which the calibrated values of the bridge condensers must be multiplied to give their absolute values, is always very near unity. If the absolute value of the capacitance of a single condenser can be determined the factor K can readily be evaluated by measuring this known condenser on the capacitance bridge. If K is known the absolute value of any condenser can then be determined by measurement on the capacitance bridge because of the consistency of the bridge calibration. The practical reason, therefore, for determining the absolute value of a primary standard of capacitance in terms of resistance and frequency is to permit the determination of the error in the bridge calibration, i.e., to evaluate K. Accordingly, the actual measurements are carried out from the viewpoint of determining the value of K for the precision capacitance bridge rather than from the viewpoint of determining the absolute value of the capacitance of a single condenser. Of course, the latter determination is included in the former.

Special Apparatus

Aside from the shielded capacitance bridge the following special apparatus employed in the determinations is worthy of mention. The unit standard condensers were dry stack mica condensers potted in an asphalt moisture-proofing compound and shielded by brass cans. They had been kept in the laboratory a number of years so that they were thoroughly aged and their values extremely stable. The phase difference of these condensers was very small, even for high grade mica condensers.

Unit resistances were made up especially for this series of tests and consisted of bifilar windings in 100 ohm sections connected in series on hard rubber spools ¾ in. in diameter. No. 40 B. & S. gauge advance wire was used throughout. All the coils had phase angles less than .1 minute at 1,000 cycles. The 6 dial shielded resistance box used in some of the measurements was a laboratory standard variable from .01 to 10,000 ohms and calibrated for phase angle. The oscillator employed as a source of current was a specially constructed vacuum tube oscillator designed to maintain an extremely constant frequency, and to deliver a practically pure sine wave. The reference standard of frequency was a 100-cycle tuning fork surrounded by a constant temperature bath,[6] the average frequency of which from day to day was constant to .001 per cent. The reference standard of resistance against which the resistances of the units were calibrated was of the well-known

[6] J. W. Horton, N. H. Ricker, W. H. Marrison, "Frequency Measurement in Electrical Communication," *A. I. E. E. Transactions*, June, 1923.

National Bureau of Standards type,[7] calibrated by the Bureau of Standards.

Experimental Procedure

The procedure used in the determination of K, was briefly as follows:

Separate unit mica condensers were selected for C and C_1. Each was measured on the capacitance bridge by itself to determine its value in terms of the bridge calibration, and in addition its series resistance was measured by comparison with the air condensers of the bridge, the series resistance of the bridge condensers being eliminated by virtue of the construction of the bridge,[5] and the method of making the measurement. At the same time the resistance of r and R, high quality resistance units, was determined by the customary Wheatstone bridge method, and their phase angles were measured by comparison with standards of which the phase angle was known to .02 minute at 1,000 cycles. The series impedance was then placed in one arm of the capacitance bridge which had previously been balanced at the frequency in question, and the parallel impedance in the other arm. The bridge was then rebalanced by varying slightly the small air condenser in the bridge and the frequency. The change in the bridge air condenser represents an algebraic addition to the capacitance C necessary because the quantities C, C_1, R, and r were not perfectly adjusted to their nominal values and because K is not exactly equal to 1. The change in frequency is necessary for the same reason. The true frequency was then determined by comparison with the laboratory standard by means of the cathode ray oscillograph.[8] For some of the determinations the bridge condensers were used for C instead of an external unit, and a shielded six dial resistance box, variable from .01 to 10,000 ohms instead of a unit resistance for R. In this case the final balance was obtained by varying the bridge condenser and R instead of the bridge condenser and the frequency. The vacuum tube oscillator used as a frequency source was capable of maintaining a frequency constant to better than .001 per cent for the duration of the tests. Sets of tests were made at three different times with an interval of about a month between them. The tests were made at frequencies of 1,000 and 2,000 cycles.

The results of the determination at each frequency are contained in Tables II and III respectively.

[7] E. B. Rosa, "A New Form of Standard Resistance," *Bulletin, Bureau of Standards*, Vol. 5, p. 413.

[8] F. J. Rasmussen, "Frequency Measurements with the Cathode Ray Oscillograph," *A. I. E. E. Journal*, January, 1927.

TABLE II

DETERMINATION OF K AT 1000 CYCLES

Those readings made on any given day are grouped together.

C' $\mu\mu f$	$(C+C')$ μf	r_1 ohm	r_1+r ohms	C_1 μf	$C_1-(C+C')$ μf	f cycles	K	$d\times 10^5$
+ 4	.100339	.15	687.08	.400130	.299791	1000.38	1.00030	+2
− 3	.100047	.15	793.53	.199133	.099086	1002.13	1.00023	−5
0	.199733	.15	397.29	.400130	.200397	1002.57	1.00025	−3
+ 4	.100336	.15	687.03	.400144	.299808	1000.49	1.00028	0
− 3	.100048	.15	793.45	.199142	.099094	1002.19	1.00028	0
0	.199725	.15	397.32	.400144	.200419	1002.61	1.00018	+10
+19	.071836	.35	998.44	.100297	.028461	1000.06	1.00031	+3
+20	.077748	.15	998.24	.199142	.121394	"	1.00036	+8
+15	.065825	.19	998.28	.301655	.235830	"	1.00032	+4
+13	.054783	.15	998.24	.400144	.345361	"	1.00036	+8
+19	.091217	.35	500.50	.100297	.009080	1000.00 ±.05	1.00032	+4
+22	.143050	.15	500.30	.199142	.056092	"	1.00031	+3
+14	.158788	.19	500.34	.301655	.142867	"	1.00024	−4
+46	.154923	.15	500.30	.400144	.246221	"	1.00023	−5
+19	.071841	.35	998.42	.100294	.028453	"	1.00023	−5
+20	.077765	.15	998.22	.199135	.121370	"	1.00025	−3
+15	.065842	.19	998.26	.301644	.235802	"	1.00019	−9
+13	.054801	.15	998.22	.400124	.345323	"	1.00029	+1
+19	.071786	.35	998.42	.100294	.028508	1001.30	1.00031	+3
+20	.077638	.15	998.22	.199131	.121493	"	1.00031	+3
+15	.065703	.18	998.25	.301641	.235938	"	1.00033	+5
+13	.054673	.15	998.22	.400122	.345449	"	1.00031	+3
+22	.142942	.15	500.35	.199131	.056189	"	1.00022	−6
+14	.158586	.18	500.30	.301641	.143055	"	1.00022	−6
+46	.154661	.15	500.35	.400122	.245461	"	1.00026	−2
+40	.100209	.15	687.03	.400122	.299913	1001.30	1.00030	+2
+19	.100134	.15	793.47	.199131	.098997	"	1.00027	−1

n = no. of observations d = deviation from mean Av. 1.00028

$\sigma = \sqrt{\frac{\Sigma d^2}{n}}$ not defined until Table 6. σ .000047 3σ .00014

Final Value of K 1.00034 ± .00003.*

To determine the effect of any inequality in the ratio arms of the bridge tests were made first with the series circuit in one arm of the bridge and then in the other at the time of each series of tests. In each case the reversal was found to cause a change of +.008 per cent in the value of K. Hence, the error in the determinations caused by all bridge inequalities was assumed as −.004 per cent; i.e., .004 per cent should be added to all determinations.

* The final value was obtained by adding .00002 for a known error in frequency and .00004 for the ratio arm error of the bridge to the average of the above determinations. The accuracy of the final value is equal to $\pm \frac{3\sigma}{\sqrt{n}}$.

TABLE III

DETERMINATION OF K AT 2,000 CYCLES

C' $\mu\mu f$	$(C+C')$ μf	r_1 ohm	r_1+r ohms	C_1 μf	$C_1-(C+C')$ μf	f Cycles	K	$d\times 10^5$
+20	.038819	.16	998.23	.100286	.061467	2000.00	1.00025	−1
+13	.027499	.06	998.13	.199108	.171609	2000.00	1.00028	+2
+14	.019687	.08	998.15	.301597	.281910	2000.00	1.00030	+4
+14	.015274	.06	998.13	.400064	.384790	2000.00	1.00025	−1
+22	.071752	.16	500.30	.100286	.028534	2000.00	1.00021	−5
+46	.077563	.06	500.20	.199108	.121545	2000.00	1.00023	−3
+43	.065626	.08	500.22	.301597	.235971	2000.00	1.00021	−5
+34	.054604	.06	500.20	.400064	.345460	2000.00	1.00025	−1
+20	.038778	.16	998.23	.100286	.061508	2001.66	1.00030	+4
+13	.027457	.06	998.13	.199116	.171659	2001.66	1.00032	+6
+22	.071712	.16	500.33	.100286	.028574	2001.66	1.00032	+6
+46	.077475	.06	500.23	.199116	.121641	2001.66	1.00025	−1
+43	.065529	.08	500.25	.301616	.236087	2001.66	1.00027	+1
+34	.054517	.06	500.23	.400082	.345565	2001.66	1.00026	0

d = deviation from mean

$\sigma = \sqrt{\frac{\Sigma d^2}{n}}$

Av. 1.00026

σ .000035

3σ .00010

Final value of K 1.00032 ± .00003.*

The resistance of the coils used for r was determined before and after each set of tests. Although the best grade of commercial resistance wire was used in making up the resistance coils they were found to have an appreciable temperature coefficient for work of such high precision, and due allowance had to be made for temperature variations in the computation of the results. The temperature coefficients of the resistances used are contained in Table IV.

TABLE IV

TEMPERATURE COEFFICIENTS OF RESISTANCE COILS USED AS r

Nominal Resistance of Coil	Temp. Coeff. % per ° C. at 20° C.
500 ohms	−.0016
400 "	+.0083
690 "	+.0013
800 "	+.0013
1000 "	+.0013

The temperature of the room in which all the tests except the measurement of resistance, were made, was held within ±1° C. As the condensers in the capacitance bridge and the special unit condensers

* The final value was obtained by adding .00002 for a known error in frequency and .00004 for the ratio arm error of the bridge to the average of the above determinations. The accuracy of the final value is equal to $\pm\frac{3\sigma}{\sqrt{n}}$.

were all so selected as to have negligible temperature coefficients over this range no temperature correction in the capacitance values nor in the value of K was necessary.

A well aged 3 dial condenser box which had just been calibrated at the Bureau of Standards was checked on the bridge at the time of the second series of tests. In this way a comparison of the true capacitance of the bridge condensers as determined by the Bureau of Standards method and as determined by the present method is afforded. This comparison is shown in Table V.

TABLE V

COMPARISON OF ACCURACY OF PRIMARY STANDARDS BY DETERMINATION OF K AND BY BUREAU OF STANDARDS CALIBRATION

	K for Bridge at 1000 cycles by method of this paper:— 2/27/26 to 3/19/26		K for Bridge at 1000 cycles by comparison with Bureau of Standards: 3/1/26
Average	1.00034	(18 values)	1.00038
σ	.000047		.00005
3σ	.00014		.00015
$\frac{3\sigma}{\sqrt{n}}$	.000027		.000035

	K for Bridge at 2000 cycles by method of this paper:— 3/19/26
Average (14 determinations	1.00032
σ	.000035
3σ	.00010
$\frac{3\sigma}{\sqrt{n}}$	.000027

Note: K by method of this paper was determined by the following bridge condensers .1, .2, .3, .4, .04, .06, .07, .08, .09. K by comparison with the Bureau of Standards values on condenser box No. 26962 was determined by all the settings of the .01 and .1 dials of the bridge.

DISCUSSION OF RESULTS

In the discussion which follows an attempt will be made to point out the various sources of error which may creep into such a determination and to state briefly what precautions were taken to guard against them.

1. *Frequency Errors*

The primary error in the values of frequency used was to be found in a variation of the standard fork from its nominal value. The average frequency of this fork integrated over 24 hours against the Arlington time signals can be held constant to $\pm .001$ per cent. While this tells us little about the fluctuations from the mean value over short periods it is at least reasonable to assume that they will be of the same general order as the variation in the average. As a check on this

assumption the frequency of a very stable vacuum tube oscillator, specially constructed, when measured against the standard fork on the cathode ray oscillograph, can be shown not to vary with respect to the standard over a period of several hours by more than $\pm.001$ per cent. Accordingly unless the standard fork and the special oscillator always fluctuate in exactly the same manner, which is very unlikely, we may conclude that both remain constant to better than $\pm.001$ per cent. The use of the cathode ray oscillograph in checking the values of the oscillator used against the standard frequency introduces no error in the determination which is appreciable from the point of view of these tests.

2. *Modulation Errors*

Another type of error due to the source of frequency used was the masking of the true balance by oscillator harmonics affecting the detector system which consisted of the double-shielded output transformer of the bridge, a vacuum tube amplifier, a telephone receiver, and the human ear. The output characteristic of each of these elements is linear, or practically so, for small loads. Overloading of any one of them, however, results in a curved output characteristic which causes an appreciable amount of modulation. This is especially true of the vacuum tube amplifier. Now in a measurement of this type the bridge is balanced for the fundamental only, since the equations of balance contain the frequency as a parameter, and any harmonics present in the input will pass through into the detector circuit practically unattenuated. If they are appreciable in magnitude the elements of the detector, particularly the amplifier, become overloaded, more harmonics are generated, these harmonics are modulated in a succeeding element of the detector, and the fundamental may appear as a modulation product even though the bridge is actually balanced. Under such conditions the fundamental tone in the receiver can be eliminated only by unbalancing the bridge slightly. It was found during the tests that if the output of the oscillator was kept small no appreciable overloading occurred in the detector circuit, but that as the oscillator output was increased the amplitude of the harmonics also increased, the detector gradually became overloaded, and the bridge balance began to change as the oscillator output was varied. This effect was eliminated by the use of a filter to keep the harmonics from the oscillator out of the detector circuit.

3. *Resistance Errors*

Errors in the determination of K due to uncertainty in resistance values arose in four ways. The first source of error due to the resistance coils lay in their temperature coefficient. This coefficient was found to

be large enough to cause serious error in some of the values, one coil in particular having an extremely large temperature variation for resistances of this type. By determining the temperature coefficient of the coils used and making appropriate temperature corrections it was possible to reduce the uncertainty due to this cause to less than .001 per cent. The second source of error in resistance values resulted from the effect of humidity on the coils. They were impregnated with shellac, which absorbed moisture and swelled sufficiently to cause changes in resistance with humidity large enough to effect the results seriously. By measuring the coils both before and after each series of tests it was possible to keep the error due to this cause below .001 per cent. The use of potted coils would undoubtedly eliminate this difficulty completely. The third resistance error present arose from uncertainties as to the series resistance of the condensers C_1. These values were determined in terms of the bridge air condensers by the step-up method, on the assumption that the air condensers in the bridge have no conductance, it being eliminated by the method of measurement and by virtue of the special construction of the bridge.

In the fourth place the accuracy with which our primary standards of resistance are calibrated by the Bureau of Standards must be considered. These calibrations have been found consistent from year to year to about $\pm.001$ per cent., and accordingly can probably be relied on to that value in the future. The results furnished by the Bureau are based on their primary standards, which agree with the primary standards of leading European nations to better than $\pm.002$ per cent, hence there is little likelihood of their changing their values by the latter amount.

4. *Capacitance Errors*

The accuracy with which the ratio of any two capacitances may be determined in such an investigation is dependent upon the consistency of the bridge calibration by the step-up method. A detailed discussion of the consistency of this calibration is beyond the scope of this paper. As an indication of the order of the precision with which it is possible to obtain a comparison between two condensers by this method, the results of measurements on a standard condenser box on three different bridges all calibrated by means of the step-up are shown in Table VI. The value of $\pm 3\sigma$, which is taken as the measure of the accuracy with which measurements can be reproduced, is $\pm.004$ per cent as determined from 54 individual measurements. On this basis the error in the ratio of two condensers compared by this method will be less than $\pm.0056$ per cent ($\sqrt{2} \times .004$). Actually the error in the comparison of the values of two condensers by measurement on such a calibrated

TABLE VI

AGREEMENT BETWEEN MEASUREMENTS ON A STANDARD CONDENSER BOX ON THREE DIFFERENT BRIDGES, ALL CALIBRATED BY THE STEP-UP METHOD, USING THE SAME PRIMARY STANDARD

Setting μf	No. 2 Bridge		No. 8 Bridge		No. 9 Bridge	
	d	$d^2 \times 10^8$	d	$d^2 \times 10^8$	d	$d^2 \times 10^8$
.01	−.0013	169	+.0007	49	+.0007	49
2	+.0003	9	+.0013	169	−.0017	289
3	+.0017	289	+.0017	289	−.0033	1089
4	.0000	0	+.0010	100	−.0010	100
5	+.0003	9	−.0007	49	+.0003	9
6	+.0014	196	−.0006	36	−.0006	36
7	+.0010	100	.0000	0	−.0010	100
8	+.0010	100	.0000	0	−.0010	100
9	+.0010	100	−.0010	100	.0000	0
.1	+.0013	169	+.0003	9	−.0017	289
2	+.0023	529	.0000	0	−.0021	442
3	+.0033	1089	−.0017	289	−.0017	289
4	+.0028	784	−.0012	144	−.0017	289
5	+.0030	900	−.0010	100	−.0020	400
6	+.0010	100	.0000	0	−.0010	100
7	+.0013	169	−.0007	49	−.0007	49
8	+.0013	169	+.0003	9	−.0017	289
9	+.0020	400	−.0005	25	−.0010	100

$$\Sigma d^2 \times 10^8 = 11215,$$

$$\sigma = \sqrt{\frac{11215}{54}} \times 10^{-4},$$

$$\sigma = .0014\%,$$

$$3\sigma = .0042\%.$$

Note: d = per cent deviation from the mean value for any setting. $\sigma = \sqrt{\frac{\Sigma d^2}{n}}$, where n = the number of observations.

bridge was probably appreciably less than ±.005 per cent in all cases, as the values of Table VI were obtained in the course of the routine calibration of the several bridges, while in the determinations of K only the one bridge was used and special precautions were taken to make the consistency of the step-up as high as possible. In any event this value of $\pm 3\sigma$ is appreciably less than that for a single determination of K, which ranges from ±.010 per cent to ±.014 per cent, as shown in Tables II and III. Accordingly the assumption that the consistency of the step-up method of calibration is sufficiently high for the purpose is well founded.

The accuracy of the capacitance values is also dependent upon the accuracy of the determination of the equivalent shunt capacitance of the resistance R. The phase angle of the series resistance r was

small enough to be neglected in all cases. The final determination of the shunt capacitance of R was accurate to $\pm 1\mu\mu f$, and hence in most of the cases under consideration the resulting uncertainty in K was less than $\pm .001$ per cent.

4. *Bridge Errors*

The principal source of error in the bridge itself lies in the inequality of the ratio arms, both in magnitude and angle. Since this inequality may result from sources other than the ratio arms proper, it is best to

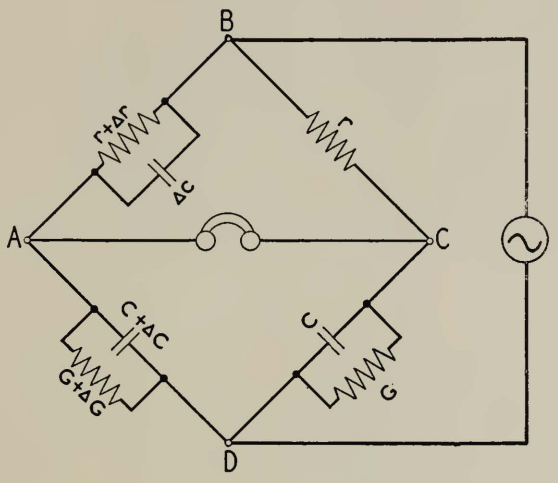


FIG. 2

ascertain it by interchanging the impedances being measured rather than by reversing the arms themselves. The following formulæ (see Fig. 2) show the errors in conductance and capacitance which may arise from ratio arm inequalities.

$$\frac{\Delta G}{2G} = -\frac{\Delta r}{r} - \frac{\omega^2 C \Delta c r}{G}, \tag{4}$$

$$\frac{\Delta C}{2C} = -\frac{\Delta r}{r} + \frac{\Delta c r G}{C}. \tag{5}$$

In the above,

G = the conductance of the unknown.

ΔG = the change in conductance due to reversing the unknown and the standard arms.

C = the capacitance of the unknown.

ΔC = the change in capacitance due to reversing the unknown and the standard arms.

r = the resistance of the ratio arms.

Δr = the resistance unbalance of the ratio arms.

Δc = the capacitance unbalance in the ratio arms.

These formulæ are not rigorous, as second order quantities have been neglected, but they are accurate to a close approximation provided the ratio arm capacitance is very small, the frequency is in the audible range, and the ratio of susceptance to conductance in the unknown is one or larger. All of these conditions obtain in the case in point.

By measuring direct and reversed an admittance having a Q (ratio of susceptance to conductance) of approximately 1 and solving the two equations simultaneously we may ascertain errors due to the differences in resistance and reactance of the ratio arms. The total change in capacitance of an admittance under test due to the resistance error of the ratio arms was found by the above method to be approximately .004 per cent; the capacitance error due to the reactance unbalance of the ratio arms was found to be $\frac{.004\%}{Q}$. Thus the combined error in capacitance due to both types of unbalance is a function of the Q of the impedance being measured. In the case of the particular type of tests being made the combined error takes the form of an error in the capacitance C, i.e., the capacitance in the shunt circuit. Table I contains a column showing the Q's of the impedances used for the tests, which range between .2 and 3.2. The corresponding error in C due to the total ratio arm unbalance varies between .014 per cent and .002 per cent (the error in C is obviously ½ of the total capacitance change resulting from the impedance arm reversal). It can easily be shown, however, from the relation between the capacitances C_1 and C of Table I that the error in K resulting from the foregoing capacitance error will in general lie between limits of .003 per cent to .005 per cent except for one or two extreme cases for which the limits are .002 per cent and .008 per cent. Accordingly the total correction due to the bridge errors was lumped at .004 as noted under "Experimental Procedure," since the few cases for which the error reaches the extreme limits are those for which the accuracy of the test as a whole is a minimum, aside from this particular type of error.

Final Accuracy of Result

The standard deviation σ for the individual determinations of K has been worked out for the values in Tables II and III. The value of σ is significant in that, provided the distribution of errors is approxi-

mately normal, over 99 per cent of all determinations will fall within limits of $\pm 3\sigma$ from the mean. In this discussion the accuracy of any measurement (exclusive of known consistent errors) will be defined as $\pm 3\sigma$. The standard deviation of the mean is given by the expression $\frac{\sigma}{\sqrt{n}}$, where n is the total number of observations. If the curve of errors is approximately normal (and we have no reason to assume otherwise), the error in the determination of K is given by $\pm \frac{3\sigma}{\sqrt{n}}$. In the absence of any systematic errors, of which none have been detected of magnitude comparable with the final accuracy of the result, the limits of accuracy are therefore, from Table V, $\pm .003$ per cent.

This limit was not exceeded in practice as is shown by the values for K at 1,000 and at 2,000 cycles in Table V. The difference between the two values of K is .002 per cent. Since the calibrations at both frequencies are based on the same original standard, namely, the 1,000-cycle value of the bridge .01μf air condenser, and the latter is assumed not to vary with frequency over the audio range, the final result for K in the two cases should be the same within the limits of accuracy of the result.

Table V contains a comparison of the values of K as determined by the method of this report and by comparing the Bureau of Standards calibrated values on a standard condenser box with the calibration of the capacitance bridge at 1,000 cycles. The agreement between these values, .004 per cent, is very close, in view of the accuracy which the Bureau certifies and the precision to which their results are given. Although the Bureau calibration is certified only to $\pm .1$ per cent, the values are furnished to 5 significant figures and are apparently consistent to $\pm .01$ per cent or better.

It will be noted that the values of capacitance chosen for these tests were all between .01 and .5μf. It is advisable that they be kept within these limits at the frequencies used in order that the resistance values required in the determination may be easily secured and easily capable of measurement with the required precision, and that errors in capacitance due to slight changes in the position of leads and units may not be appreciable.

Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks

By OTTO J. ZOBEL

Synopsis: Constant resistance recurrent networks, that is, networks whose iterative impedances are a pure constant resistance at all frequencies, form here the basis of a method of distortion correction which is applicable to any electrical circuit. The paper takes up first the general problem of distortion correction, then this method of correction and its application in the following Parts and supplementary Appendices.

Part 1. *Ideal Circuit Characteristics.* Both ideal steady-state attenuation and phase characteristics are formulated and then verified as being necessary and sufficient for the preservation of signal-shape under transient conditions.

Part 2. *Constant Resistance Recurrent Networks.* These networks are of three general types and are made possible by the introduction of inverse networks of constant impedance product. Their propagation characteristics are considered in some detail and various methods of design are indicated.

Part 3. *Arbitrary Impedance Recurrent Networks.* These networks are a generalization of those in Part 2.

Part 4. *Applications.* The large variety of uses to which these networks may be put is illustrated by specific designs made for complementary distortion correcting networks, for a submarine cable circuit, a loaded-cable program transmission circuit, and an open-wire television circuit. In addition, networks are given for the equalization of variable attenuation in carrier telephone circuits, for phase correction in the transatlantic telephone system and for the simulation of a smooth line.

Appendix I. *Discussion of Linear Phase Intercept.*

Appendix II. *Linear Transducer Theorems.*

Three theorems are proved which relate to the variation with frequency over the entire frequency range of the propagation constants and iterative impedances of certain passive linear transducers.

Appendix III. *Propagation Constant and Iterative Impedance Formulæ for General Ladder, Lattice and Bridged-T Types.* This includes an improved formula for $\cosh^{-1}(x + iy)$.

Appendix IV. *Propagation Characteristics and Formulæ for Various Lattice Type Networks.* These results can be applied quite readily to many problems arising in the design of distortion correcting networks.

Introduction

EVERY actual electrical circuit or transmission system distorts transmitted signals; that is to say, the received signal, regarded as a time-function, differs in shape from the impressed signal. Heaviside studied in detail the distorting action of the transmission line itself and indicated the necessary electrical properties of the distortionless line.[1] The distortionless line of Heaviside was approximately realized in the loaded line [2] in which similar lumped inductances are inserted in series with the line at uniform intervals. While this loading has the effect of partially correcting distortion of the lower frequency components of the signal, it also tends to increase the distortion of the higher frequency components and so limit somewhat the useful frequency range. More recently the transmission characteristics of some newly installed submarine cables have been greatly improved by means of continuous loading with the new magnetic material permalloy.[3]

The methods mentioned above are directed to rendering the line itself more nearly perfect. The method of distortion correction presented here may be used to supplement them and is that of passive *terminal* networks; more particularly networks whose iterative impedances are a pure constant resistance at all frequencies.[4] These networks are, however, not limited in their use to any particular type of transducer or transmission system but have general applicability. For this reason the general problem of distortion correction by this method resolves itself principally into a study of the transmission properties of these networks together with systematic methods of design to meet specified requirements.

This paper takes up first the characteristics necessary for no distortion in an electrical circuit; then, an extended study of constant resistance networks which can be used for distortion correction; finally, several applications to important practical problems. In addition, Appendix IV gives a considerable number of network structures and

[1] "Electrical Papers," Vol. II, p. 123, 1892; "Electromagnetic Theory," Vol. I, p. 445, 1893, Oliver Heaviside.

[2] U. S. Patent No. 652,230 to M. I. Pupin, dated June 19, 1900. See also "On Loaded Lines in Telephonic Transmission," G. A. Campbell, *Phil. Mag.*, March, 1903. Later a loading system more specifically directed to reducing distortion *per se* was disclosed in U. S. Patent No. 1,564,201 to J. R. Carson, A. B. Clark and J. Mills, dated December 8, 1925.

[3] "The Loaded Submarine Telegraph Cable," O. E. Buckley, *B. S. T. J.*, July, 1925.

[4] The equalization of the attenuation of certain transmission lines has for some time been obtained by means of comparatively simple series or shunt terminal networks. See, for example, U. S. Patent No. 1,453,980 to R. S. Hoyt, dated May 1, 1923. Such networks necessarily produce total terminal impedances which vary with frequency.

corresponding formulæ which will be found useful in further applications.

Part 1. Ideal Circuit Characteristics

There is no distortion in the transmission of an impressed signal over an electrical circuit or network when the shape of the received signal, considered as a time-function with usually a time-of-transmission, is identical with that of the impressed signal. A uniform decrease in magnitude only is not distortion, and it can be restored to its original value by means of a distortionless amplifier.

Let us assume in the general case that the e.m.f. impressed on the circuit is E, and that the circuit is always terminated by a receiver of resistance, R, across which is the received voltage, v, in which we are interested. The received current is then directly proportional to the received voltage.

The necessary and sufficient conditions for distortionless transmission can be stated quite simply in terms of the steady-periodic transfer voltage ratio of the circuit which will be written as

$$\frac{v(i\omega)}{E(i\omega)} = e^{-a-ib}, \tag{1}$$

with the terminology $a + ib$ = *the transfer voltage exponent of the circuit*, or concisely, *the transfer exponent*. Here a represents attenuation in napiers and b phase difference in radians, omitting in the latter any constant integral multiple of 2π, and assuming the two voltages to have zero phase difference at zero frequency. That is, the origin of phase difference is so chosen that the phase intercept at zero frequency is zero.

For ideal transmission characteristics the steady-periodic transfer exponent of the circuit should have an attenuation independent of frequency and a phase proportional to angular frequency, ω, whose slope is the time-of-transmission of the circuit.

In mathematical terms these ideal characteristics, represented by primes, are

$$\begin{aligned} a' &= \text{constant (napiers)}, \\ \text{and} \qquad & \qquad (2) \\ b' &= \tau\omega \text{ (radians)}, \end{aligned}$$

where

$$\tau = \text{time-of-transmission (seconds)}.$$

To show this, consider first what the indicial voltage, $g(t)$, would be under these assumptions. By *indicial voltage* is meant the received voltage as a time-function per unit constant e.m.f. impressed at the

sending end at time $t = 0$. With (1) and (2) in the integral equation of electric circuit theory [5] we obtain

$$\frac{e^{-a'-\tau p}}{p} = \int_0^\infty e^{-pt} g(t)dt, \tag{3}$$

whose solution is

$$g(t) = 0, \quad t < \tau,$$

and

$$g(t) = e^{-a'} = \text{constant}, \quad t > \tau. \tag{4}$$

Thus, a constant voltage, which has been attenuated by the circuit an amount a' napiers, arrives suddenly at the receiving end after a time $\tau = (b'/\omega)$ seconds, and there is no distortion with respect to the unit constant e.m.f. impressed on the circuit at time $t = 0$.

If now any type of e.m.f., $E(t)$, is impressed on this circuit which is specified by the steady-state characteristics (2) or the indicial voltage (4), we obtain through a general formula [6]

$$v(t) = \frac{d}{dt}\int_0^t E(t - y)g(y)dy = e^{-a'}E(t - \tau). \tag{5}$$

This received voltage has the same shape as the impressed e.m.f., there being an attenuation, a', and a time-of-transmission, τ. Hence, *a circuit specified as above is distortionless to any type of impressed e.m.f.* A further discussion involving the phase intercept is taken up in Appendix I.

It may be stated that Heaviside's theoretical distortionless smooth line was that in which the line constants R', L', G' and C' per unit length had the relation

$$R'/G' = L'/C', \tag{6}$$

giving attenuation and phase constants per unit length, respectively,

$$\alpha = \sqrt{R'G'} \text{ napiers},$$

and

$$\beta = \sqrt{L'C'}\omega \text{ radians};$$

also an iterative (or characteristic) impedance

$$k = \sqrt{\frac{R'}{G'}} = \sqrt{\frac{L'}{C'}} \text{ ohms},$$

which is a constant resistance at all frequencies. A circuit made up

[5] "Electric Circuit Theory and the Operational Calculus," John R. Carson.
[6] L.c.

of such a line of length l terminated by a resistance $R = k$ is readily seen to satisfy the conditions (2) above for no distortion. It would have an attenuation $a' = \sqrt{R'G'}l$ napiers and time-of-transmission $\tau = \sqrt{L'C'}l$ seconds.

Having seen above what constitutes ideal transmission characteristics, the problem of distortion correction in any practical distorting circuit is that of altering the circuit in some way so as to approach this ideal. In most circuits it is impossible to obtain these ideal characteristics throughout the entire frequency range. More or less satisfactory transmission results will be had, however, if this ideal is approached over the range of frequencies most essential to the composition of the impressed e.m.f., as shown by its Fourier integral analysis.

How accurately an ideal attenuation characteristic has been met in any case depends upon how nearly constant the attenuation is in the frequency range. A simple practical measure of the degree of approach to an ideal phase characteristic at the frequencies in this range is furnished by a consideration of the time-of-phase-transmission in the steady state,

$$\tau_p = b/\omega \text{ seconds}, \tag{7}$$

in which b is defined as in (1) for the complete circuit. The more nearly constant τ_p is in the frequency range, the closer it approaches equality with τ, the time-of-transmission of the circuit for those frequencies.

In many cases approximately ideal phase characteristics already exist in the desired frequency ranges so that corrections need be made for attenuation only. In others, such as those in which the steady-periodic state is of most importance and where the phase relations between the components are immaterial, it is satisfactory to obtain uniform attenuation at the desired frequencies. The method of altering circuit transmission characteristics to be shown in this paper follows in Part 2.

Part 2. Constant Resistance Recurrent Networks

2.1. *Fundamental Basis of Distortion Correction*

The general transmission circuit of Fig. 1 is shown as having a resistance, R, at the receiving end, as in the case where the energy is absorbed. Usually the circuit characteristics at this resistance with respect to the sending terminals show distortion in the required frequency range. If so, an ideal method of correcting the distortion

would appear to be that of interposing between the circuit and the receiving resistance a transducer having the requisite corrective propagation constant and an iterative impedance, R. By so doing, the transfer exponent at the end of the circuit proper would remain un-

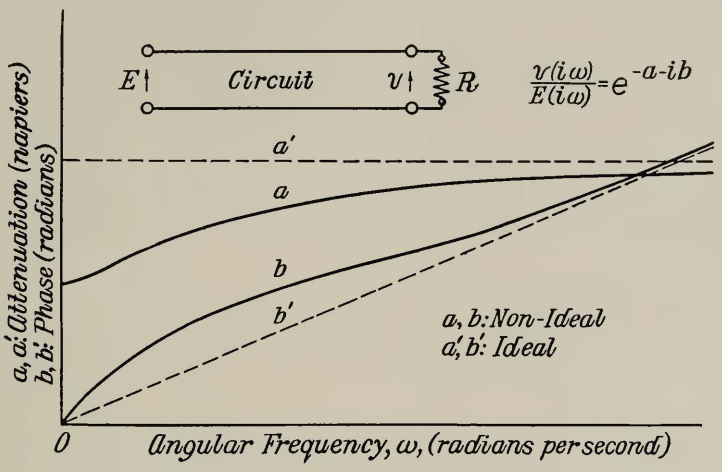


Fig. 1—Non-ideal and ideal transfer exponents of circuits.

altered, *irrespective of the exact nature of the network beyond*, since the latter has the impedance R; but the total transfer exponent would become ideal through the addition of the complementary propagation constant of the transducer. Stated analytically,

let $a + ib$ = transfer exponent of the distorting circuit at a terminating resistance R,

$A + iB$ = propagation constant of the correcting transducer of iterative impedance R,

and $a' + ib'$ = resultant ideal transfer exponent at the receiving resistance R.

Then the correcting transducer must be so designed that A and B satisfy over the required frequency range the conditions

$$a' = a + A = \text{constant},$$

and

$$b' = b + B = \tau\omega,$$

where τ is a positive constant. Or, explicitly,

$$A = a' - a, \text{ positive},$$

and (8)

$$B = b' - b = \tau\omega - b.$$

The total attenuation, a', and time-of-transmission, τ, are somewhat at our disposal; it will be found that their best choice is usually guided by experience. The transducer will often consist of a number of sections, not necessarily alike. This distortion correcting process may be called "equalizing both the attenuation and the time-of-phase-transmission."

The idea of altering circuit transmission characteristics by means of one or more sections of constant resistance recurrent networks forms the fundamental basis of the method of distortion correction presented here. It is, of course, dependent for its application upon the physical possibility of designing recurrent networks whose iterative impedances are a constant resistance at all frequencies and whose propagation constants have the desired characteristics.

Another method by which distortion correction has sometimes been obtained is by means of terminal thermionic distortion circuits wherein networks of particular frequency characteristics are placed in the plate circuits of successive thermionic tubes. In it any reaction of one stage upon a preceding stage or upon the original circuit is prevented by the unilateral property of the tubes, whereas in the method given here this same result is obtained by the property of a constant resistance iterative impedance and the use of a resistance termination. While from the standpoint of the original circuit both methods give the resultant effect of a terminal unilateral device, one very practical advantage of the constant resistance method over the thermionic tube method appears to be that it corrects distortion before any amplification is added and hence with it there would be less tendency to cause tube distortion or modulation. Another advantage is that the distortion correcting networks can be designed independently of the amplifying device. A description of this other method appeared in the last number of the *Journal*.[7]

Before taking up specific types of constant resistance structures, let us consider some of the inherent limitations of certain transducers as are brought out by the following theorems.

2.2. *Linear Transducer Theorems*

These theorems relate to the variation with frequency over the entire frequency range of the iterative parameters, that is, the propagation constants and iterative impedances, of certain passive linear transducers. In symmetrical transducers we could as well employ the image parameters which are of such utility in a study of electric wave-

[7] "Phase Distortion and Phase Distortion Correction," Sallie Pero Mead, *B. S. T. J.*, April, 1928.

filters and which, together with iterative parameters, were discussed generally by the writer in a previous number of this *Journal*.[8] But since here in the ladder type networks some dissymmetrical sections are also considered, I shall use the iterative parameters throughout this paper.

Theorem I: Any symmetrical transducer whose attenuation constant is zero at all frequencies has a phase constant which increases with frequency and an iterative impedance which is a constant resistance throughout the frequency range.

Theorem II: Any transducer whose iterative impedance is real at all frequencies has a constant resistance iterative impedance, and if in addition its phase constant is proportional to frequency, it has a uniform attenuation constant.

Theorem III: Any symmetrical transducer whose attenuation constant is independent of frequency and whose iterative impedance is a constant resistance at all frequencies has a phase constant which is zero or increases with frequency.

The theorems, whose proofs are given in Appendix II, may be represented by the following table. The variations with frequency of the network parameters shown apply to the entire frequency range and in each theorem the parenthesis designates the dependent property, where A is the attenuation constant, B the phase constant, and K the iterative impedance.

TABLE I

LINEAR TRANSDUCER THEOREMS

Theorem	A	B	K
I.....	0	(Increases)	(Constant)
II.....	(Constant)	$\tau\omega$	Real (Constant)
III.....	Constant	(Zero, or increases)	Constant

That part of Theorem I which relates to the iterative impedance explains why there is no physical ladder type network having zero attenuation throughout the frequency range. For, the ladder type, when non-dissipative and having zero attenuation, requires a mid-series or mid-shunt iterative impedance which varies with frequency.

[8] "Transmission Characteristics of Electric Wave-Filters," O. J. Zobel, *B. S. T. J.*, October, 1924. The term "characteristic impedance" used in that paper for a recurrent or iterative parameter with dissymmetrical transducers is replaced here by "iterative impedance." Thus, the same term "iterative" applies to the structure, to the corresponding impedances, and to the kind of parameters. The use of the term "characteristic impedance" will be limited to smooth lines, or sometimes to symmetrical recurrent structures. In symmetrical structures the "characteristic," "iterative," and "image" impedances are identical.

2.3. *Inverse Networks of Constant Impedance Product*

We have already seen that the fundamental advantage of using constant resistance networks for distortion correction lies in the fact that when they are placed ahead of the receiving resistance, R, they present this same impedance to the circuit proper and hence do not alter the transfer exponent at that point. They can be designed to have, in addition to the impedance R, a propagation constant which complements this exponent and produces a resultant transfer exponent at the receiving resistance which is approximately ideal.

The possibility of physically realizing recurrent networks having a constant resistance iterative impedance at all frequencies rests, as will be seen, upon that of obtaining pairs of two-terminal networks the product of whose impedances is constant, independent of frequency. Such pairs [9] I have defined as *inverse networks of impedance product* R^2, or more concisely, *inverse networks.*

In the paper just referred to it was pointed out that one elemental pair of such inverse networks is composed of two resistances R_1 and R_2, and another is composed of an inductance L and a capacity C bearing the impedance product relations at all frequencies

$$R_1R_2 = L/C = R^2. \tag{9}$$

The same paper gave a simple proof of the following theorem relating to series and parallel combinations of networks. *If* z_1' *and* z_2' *are any pair of inverse networks and if* z_1'' *and* z_2'' *are any other pair, such that* $z_1'z_2' = z_1''z_2'' = R^2$, *then* z_1' *and* z_1'' *in series and* z_2' *and* z_2'' *in parallel are a pair; similarly* z_1' *and* z_1'' *in parallel and* z_2' *and* z_2'' *in series are another pair.*

Without much difficulty a theorem relating to simple networks having the form of a general Wheatstone bridge can also be obtained, as follows: *The inverse network corresponding to any given two-terminal bridge network of five distinct branches is also a bridge network, and may be derived by replacing the network in each branch of the given network by its inverse network and then interchanging the networks in either opposite pair of branches.* By successive applications of these relations, beginning with the elemental pairs, very complicated inverse networks can be built up. Only reactance networks were considered in the paper referred to above. Ordinarily the series and parallel

[9] An extensive use of inverse networks of pure reactance types was made in the paper, "Theory and Design of Uniform and Composite Electric Wave-Filters," O. J. Zobel, *B. S. T. J.*, January, 1923. Also in U. S. Patents No. 1,509,184, September 23, 1924; Nos. 1,557,229 and 1,557,230, October 13, 1925; and No. 1,644,004, October 4, 1927.

combinations are most useful, since the bridge structures require at least five elements in each network. Some networks may have other equivalent structures, as well.

If $z_{11} = r_{11} + ix_{11}$ and $z_{21} = r_{21} + ix_{21}$ are inverse networks such that

$$z_{11}z_{21} = R^2, \tag{10}$$

a number of simple relations exist among their impedance components; namely,

$$\frac{x_{21}}{r_{21}} = -\frac{x_{11}}{r_{11}},$$

$$\frac{r_{21}}{|z_{21}|^2} = \frac{r_{11}}{R^2}, \tag{11}$$

and

$$\frac{x_{21}}{|z_{21}|^2} = -\frac{x_{11}}{R^2}.$$

In a smooth line the condition (6) which makes it distortionless is actually the one making the series and shunt impedances per unit length inverse networks of impedance product $R^2 = R'/G' = L'/C'$.

2.4. *Types of Constant Resistance Recurrent Networks and Their Propagation Constants*

The types of recurrent networks considered in this paper are the three simplest ones, the ladder, lattice, and bridged-T types whose general structures are shown in Fig. 2. Propagation constant and iterative impedance formulæ for these types in terms of general impedance elements are given in Appendix III for possible future reference.

By introducing in each of these types the use of inverse networks with z_{11} and z_{21} satisfying relation (10), and assuming various relations in the general formulæ, it is possible to derive general network structures whose iterative impedances are a constant resistance, R, at all frequencies.[10] The structures are of such general nature as to permit a very wide range of propagation constants. Any one of them when closed by a resistance, R, presents at the other terminals the impedance R at all frequencies. They will now be considered.

The networks of the *ladder type* are shown in Fig. 3 as six complete sections, each designated by the termination at which it has the iterative impedance R; one at full-series, one at full-shunt, and two

[10] See U. S. Patent No. 1,603,305 to O. J. Zobel, dated October 19, 1926. Also British Patent Specification No. 236,189, dated July 8, 1926.

each at mid-series and mid-shunt. The first two sections are dissymmetrical as regards the two pairs of terminals. The two mid-series sections are symmetrical and identical except for the structure of their shunt branches, which, however, are equivalent impedances. Simi-

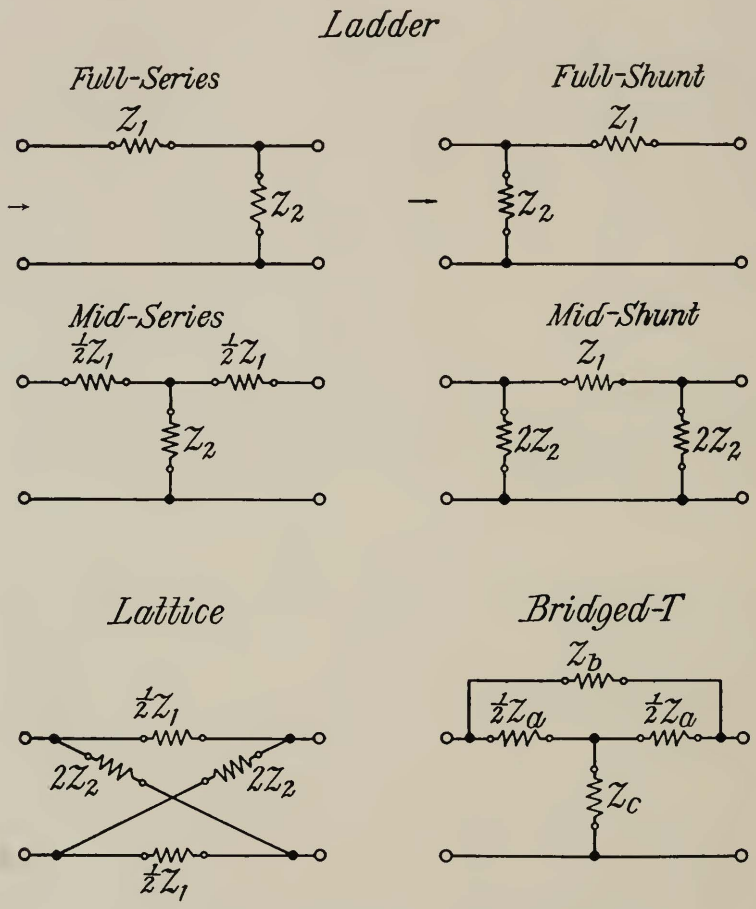


Fig. 2—Types of general recurrent network sections.

larly, the symmetrical mid-shunt pair have different series branches of equivalent impedance. It may be of interest to point out that if each of these sections is closed by a resistance, R, to form a two-terminal network, then three pairs of these networks are seen directly from series and parallel rules to be inverse networks; namely,

$$I_a,\ I_b;\ I_c,\ I_f;\ \text{and}\ I_d,\ I_e.$$

If one has the impedance R, the other must also, as is the case. The propagation constant, $\Gamma = A + iB$, of each of these *ladder type* sections is the same and is given by the simple relation

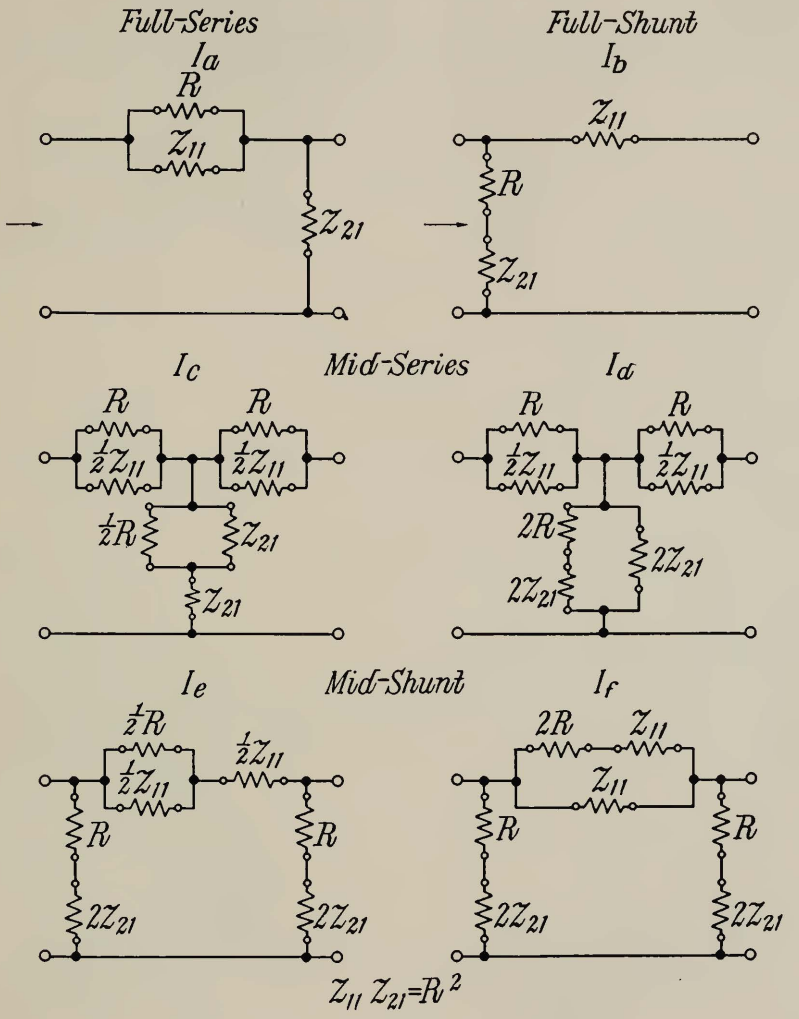


Fig. 3—Ladder type constant resistance sections.

$$e^{\Gamma} = 1 + z_{11}/R; \tag{12}$$

the particular iterative impedance is R. Here z_{11} is arbitrarily taken as the independent impedance determining the propagation constant

with z_{21} dependent through the inverse network relation $z_{11}z_{21} = R^2$. This relation, besides ensuring a constant resistance iterative impedance, reduces the network parameters at least one half. *Since resistances occur explicitly in the structures, these sections will all be dissipative.*

The network of the *lattice type*, shown in Fig. 4, is symmetrical and has a propagation constant determined by

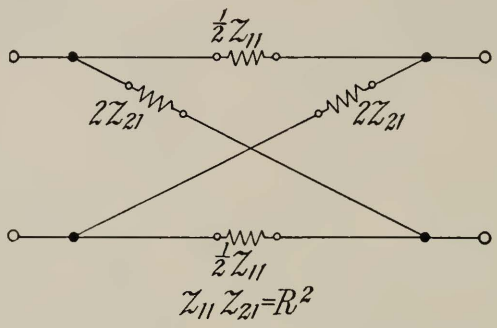


Fig. 4—Lattice type constant resistance section.

$$e^{\Gamma} = \frac{1 + z_{11}/2R}{1 - z_{11}/2R}, \tag{13}$$

where $z_{11}z_{21} = R^2$. *If z_{11} is a reactance, the network will introduce no attenuation, only phase difference.*

The networks of the *bridged-T type* are symmetrical and will be given in two groups, the members of each group having the same propagation constant. The two sections of the *first group* (I_a and I_b) in Fig. 5

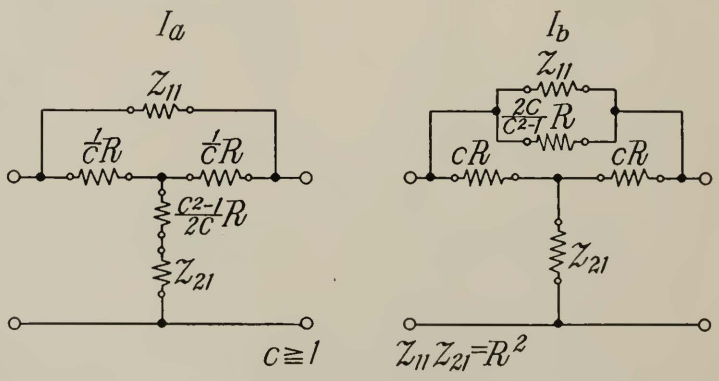


Fig. 5—Bridged-T(I) type constant resistance sections.

have a propagation constant formula

$$e^{\Gamma} = \frac{1 + (c+1)z_{11}/2R}{1 + (c-1)z_{11}/2R}, \tag{14}$$

where, besides the arbitrary impedance z_{11}, there is the arbitrary real $c \geqq 1$. *These sections will be dissipative owing to the ever present resistances.* Utilizing directly the rule given for inverse bridge networks, it can be seen that when closed by R these two structures are inverse networks of impedance product R^2.

The four bridged-T sections of the *second group* (II_a, II_b, II_c, and II_d) in Fig. 6 have the formula

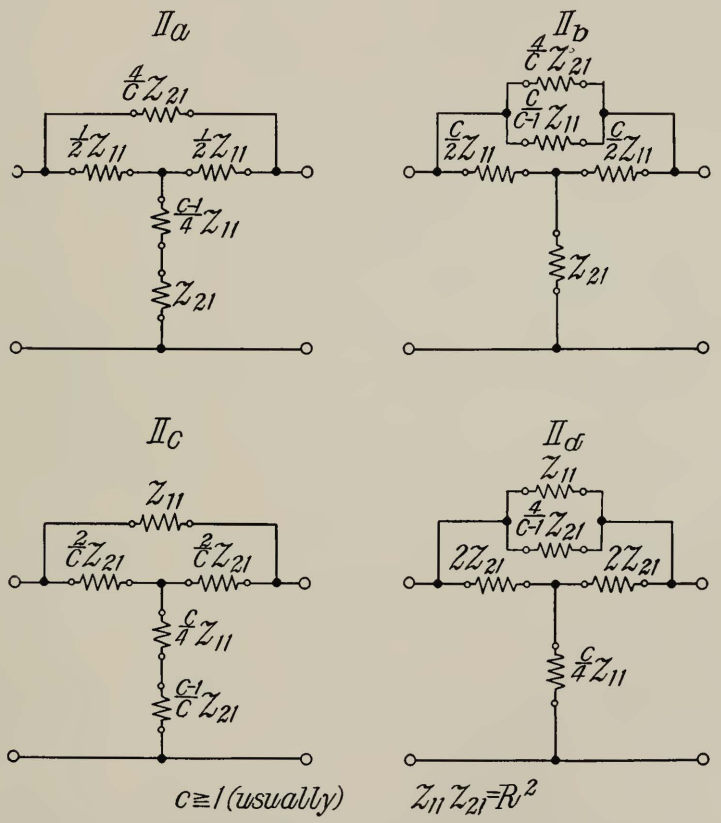


Fig. 6—Bridged-T(II) type constant resistance sections.

$$e^{\Gamma} = \frac{1 + z_{11}/2R + c(z_{11}/2R)^2}{1 - z_{11}/2R + c(z_{11}/2R)^2}, \tag{15}$$

where $c \geqq 1$, usually. In the very special cases of networks II_a and II_b, wherein z_{11} is an inductance, c may be less than unity and approach zero as a limit. For the latter values the negative inductance may be obtained physically as a negative mutual between the series coils. When $c = 1$, networks II_a and II_b become physically identical, as do also networks II_c and II_d. *If z_{11} is a reactance, there will be no attenuation.* Again, we shall find by applying the proper rule directly that when the four general sections are closed by resistances R there will result two pairs of inverse networks of impedance product R^2, respectively II_a, II_d and II_b, II_c.

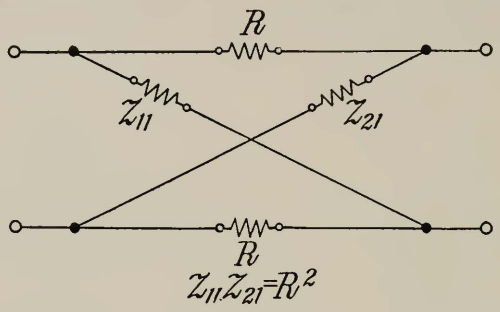


Fig. 7—Unbalanced lattice type constant resistance section.

A special network of the *unbalanced lattice type* may be mentioned briefly. This symmetrical structure as shown in Fig. 7 plays no direct part here as a distortion correcting network but is closely related to some of the other types and possesses interesting properties, among others that of conjugacy as in an ordinary balanced Wheatstone bridge. Its open-circuit impedance X and short-circuit impedance Y are both equal to R, hence its iterative impedance, $\sqrt{XY}$, is also R. Since $\tanh \Gamma = \sqrt{Y/X} = 1$, $\Gamma = \infty$, which means that no current would flow in a terminating resistance R due to an e.m.f. applied through a sending resistance R, these two impedance branches being conjugate. The network containing four resistances R, which is obtained by terminating this section at each end by a resistance R, may likewise be derived directly from the limiting case ($c = 1$) of the bridged-T (I) section which has similarly been terminated, merely by a rearrangement of form. It has these properties:

1. Opposite resistances are in conjugate branches.
2. Each of the four resistances is faced by a resistance R.

These properties can be seen as a result of the symmetry and also from a comparison with the full-series and full-shunt ladder type sections when terminated by resistances R. It is known that for one direction

of propagation these two ladder sections have the same iterative impedance and propagation constant. In the full-series section terminated by R the junction point between R of the section and z_{21} is short-circuited with the point at the receiving side of z_{11}, while in the corresponding full-shunt network the structure is the same except that these two points are open-circuited. Because of the identity of propagation constants this can be possible only if the two points are at the same potential whence they can be connected by any impedance without altering propagation in the one direction. This being the case, a branch of resistance R, conjugate with the sending branch, can be connected across these points, and this results in giving the symmetrical bridged-T (I) type (where $c = 1$), or the equivalent network of Fig. 7 terminated by R. Thus the receiving-side series resistance R in the limiting case ($c = 1$) of the bridged-T (I) section plays no rôle and is superfluous for this direction of transmission, but it makes the section symmetrical and ensures similar propagation and impedance characteristics when transmitting in the opposite direction.[11]

If, in the network of Fig. 7, z_{11} is made resonant and anti-resonant at different frequencies, selective *maximum* energy transmission can be obtained at these frequencies between pairs of the four different resistance branches which might also be considered as different lines. The propagation constant between any pair of resistances can be determined from the relationships established above.

As an aid in obtaining an approximate value of the propagation constant for any of these types when its impedance elements are known, a simple chart may be drawn up if desired. This could be obtained in the following manner. The formulæ (12) to (15) are all of the form

$$e^{\Gamma} = e^{A+iB} = m + in;$$

whence

$$e^{A} = \sqrt{m^2 + n^2}, \tag{16}$$

and

$$\tan B = n/m.$$

Thus, it is evident that any locus of uniform attenuation constant, A, is represented in the m, n plane by a circle of radius, e^A, with center at the origin. Also, any locus of uniform phase constant, B, is a straight line of slope, tan B, starting from the origin.

[11] Another method of deriving the section having directly the form given by putting $c = 1$ in the bridged-T (I) type was used by G. H. Stevenson, U. S. Patent No. 1,606,817, November 16, 1926.

2.5. *Relations for Equivalence of Propagation Constants*

All of the above networks have equivalent iterative impedances equal to R. It is sometimes useful to be able to transform readily from one type to another which has also an equivalent propagation constant, if that is physically possible. This may arise in an economic study of a final network design where account is taken of all practical factors, such as symmetry, line balance, number of the elements, their magnitudes, etc.

The structures which are important in this connection when dealing *with both attenuation and phase characteristics* comprise the ladder, lattice, and bridged-T (I) networks, whose propagation constant formulæ are given in (12), (13), and (14). For their propagation constants to be identical the impedance z_{11} in one type must bear a definite relation to that in another. In the following table, derived by equating these formulæ, a general impedance z is introduced. Each z_{11} may be expressed in terms of z and R. Here z is taken as the z_{11} for each type in succession. It then becomes a simple matter to transform from one type of structure to another having an equivalent propagation constant. The parameter c in a derived bridged-T (I) network would be taken such as to give the minimum number of elements.

TABLE II

RELATIONS FOR EQUIVALENCE

Ladder z_{11}	Lattice z_{11}	Bridged-T (I) z_{11}, $c \geqq 1$
z	$\dfrac{1}{\frac{1}{z}+\frac{1}{2R}}$	$\dfrac{1}{\frac{1}{z}+\frac{1}{-2R/(c-1)}}$
$\dfrac{1}{\frac{1}{z}+\frac{1}{-2R}}$	z	$\dfrac{1}{\frac{1}{z}+\frac{1}{-2R/c}}$
$\dfrac{1}{\frac{1}{z}+\frac{1}{2R/(c-1)}}$	$\dfrac{1}{\frac{1}{z}+\frac{1}{2R/c}}$	z

A transformation from the z_{11} of one type section to that of another equivalent one involves essentially only an alteration of the given impedance by a positive or negative resistance element in parallel with it. This will not always result in a physical network with

positive elements. The following statements can be made, however:

1. The transformation of the ladder type to the equivalent bridged-T (I) type, and vice versa, is always possible.

2. The transformation of the ladder type, or the bridged-T (I) type, to the equivalent lattice type is always physically possible; the converse is not necessarily so.

Those structures which are potentially phase networks, and thus useful when requiring a non-attenuating network *with a phase characteristic only*, are the lattice type again and the bridged-T (II) type. Such networks are used to introduce various characteristics for the time-of-phase-transmission. It will be sufficient to give the relations for equivalence between these two types, obtained from (13) and (15), as

$$(z_{11})_{\text{lattice}} = \left(\frac{1}{\frac{1}{z_{11}} + \frac{1}{4z_{21}/c}} \right)_{\text{bridged-T (II)}} , \tag{17}$$

which is always physically possible if the bridged-T (II) network exists. On the other hand

$$(z_{11})_{\text{bridged-T (II)}} = \frac{2}{c}(z_{21} \pm \sqrt{z_{21}^2 - cR^2})_{\text{lattice}} , \tag{18}$$

where the c which belongs to the bridged-T (II) type must necessarily be taken so as to make the radical a perfect square, *if a physical equivalent is possible.* It is to be pointed out that in the propagation constant formula (15), considered as a general form, the range of values for the parameter c which will give a physical bridged-T (II) network is $c \geqq 1$, usually, while the range for a physical lattice network is $c \geqq 0$, as seen from (17). Thus, the lattice type can give a greater variety of propagation constants.

From all the comparisons made above this conclusion may be drawn. *The lattice type has a greater range for its propagation constant characteristic than has either a ladder or a bridged-T type. Hence, the lattice type might well be considered as the fundamental one, when designing such networks, from which other equivalent types may be obtained by transformations, if such physical structures are possible.*

2.6. *Propagation Constants Expressed as Frequency Functions*

In Section 2.4 the propagation constant of any of these networks was given as varying with frequency only implicitly, according to some function of the impedance ratio, $z_{11}/2R$. To express it more explicitly as a frequency function, I shall sketch briefly a satisfactory general method to be followed.

For an impedance z_{11} which is made up of lumped elements of resistance, inductance, and capacity we may express the impedance ratio $z_{11}/2R$ as the ratio of two frequency-polynomials in (if), where $i = \sqrt{-1}$ and f is frequency. Thus,

$$\frac{z_{11}}{2R} = \frac{a_0 + a_1(if) + a_2(if)^2 + \cdots}{b_0 + b_1(if) + b_2(if)^2 + \cdots} = s + iy. \tag{19}$$

The impedance coefficients a_0, b_0, etc., of which one is unity and some may be zero, are positive quantities and are algebraic combinations of the network elements. Their number is equal to, or greater than, the number of independent elements. For any given type of network the coefficients are fixed by the elements, and vice versa.

Putting this expression in any of the formulæ (12) to (15), there results *for the propagation constant* a form

$$e^{\Gamma} = \frac{g_0 + g_1(if) + g_2(if)^2 + \cdots}{h_0 + h_1(if) + h_2(if)^2 + \cdots}, \tag{20}$$

in which g_0, h_0, etc., are algebraic functions of a_0, b_0, etc., also of c if the network is a bridged-T type. From this the attenuation constant and phase constant can also be derived and expressed separately as functions of frequency.

For the attenuation constant, a form is obtained

$$F \equiv e^{2A} = 10^{TU/10} = \frac{P_0 + P_2f^2 + \cdots}{Q_0 + Q_2f^2 + \cdots}, \tag{21}$$

which is the ratio of two frequency-polynomials both in even powers of frequency. One of the attenuation coefficients is unity.

For the phase constant, a form

$$H \equiv \tan B = \frac{M_1f + M_3f^3 + \cdots}{N_0 + N_2f^2 + \cdots}, \tag{22}$$

in which one of the phase coefficients is unity, is the ratio of two frequency-polynomials, odd powers of frequency in the numerator and even powers in the denominator. (It is sometimes convenient to use tan $(B/2)$.) In (21) and (22) the attenuation coefficients P_0, Q_0, etc., and the phase coefficients M_1, N_0, etc., are expressible in terms of the impedance coefficients a_0, b_0, etc.

It should be mentioned here that in deriving the above expressions certain assumptions have been made; namely, invariable elements and non-dissipative inductances and capacities. These restrictions are well justified from the fact that such departures are usually small

and their effects in a network do not alter appreciably the general characteristics. However, to calculate accurate results for both the propagation constant and the iterative impedance of the final design of a physical network taking into account all factors, one should use the general formulæ given in Appendix III which have been simplified to give accurate results quite readily.

2.7. *Network Solutions from Their Propagation Characteristics*

It was assumed in the previous section that the recurrent network elements are invariable and that inductances and capacities are non-dissipative. On this basis general formulæ for the propagation characteristic were obtained in terms of these elements. The same assumptions are retained here but *reverse processes* will be carried through which derive the elements from the propagation characteristic of the recurrent network. Three methods will be outlined, necessarily in general terms.

Method 1. *Solutions from the Attenuation Constant*

Since attenuation is ordinarily of greatest importance, this method is the one most frequently used with networks having an attenuation characteristic and involves initially the determination of the attenuation coefficients P_0, Q_0, etc., from this characteristic. Using these coefficients, one derives from algebraic relations, first, the impedance coefficients a_0, b_0, etc., and finally the network elements in z_{11}. The elements of z_{21} follow from the inverse network relation (10).

The method is based upon the transformation of the attenuation formula (21) to a linear equation in P_0, Q_0, etc., whose number is equal to or greater than the number of independent network parameters. If we multiply equation (21) by the Q-polynomial, we obtain formally the *attenuation linear equation* which holds at all frequencies,

$$P_0 + f^2P_2 + \cdots - FQ_0 - f^2FQ_2 - \cdots = 0. \tag{23}$$

Introducing in this the attenuation constant, and hence F, at a number of different frequencies equal to the number of independent network parameters, there results a system of independent simultaneous linear equations which can be solved for the coefficients. The simplest practical procedure is perhaps that of the *step-by-step elimination of the coefficients.*

When the number of coefficients and independent network parameters, hence equations, are the same, the solution of the latter offers no particular difficulty and results can readily be checked by substitution in the original equation (21).

When, as sometimes occurs, the number of coefficients is one greater than the number of independent network parameters, it means that one relation exists between the coefficients and hence any one of the latter may be assumed dependent. The dependent relation can be found from the formulæ for P_0, Q_0, etc., in terms of a_0, b_0, etc. However, in some such networks it is possible to use the attenuation constant at a particular frequency, say zero or infinite frequency, and thereby reduce the number of remaining coefficients and independent network parameters to equality, when the case is readily solvable. If this does not produce the desired reduction, it is usually best to first transfer the dependent coefficient to the right-hand member of (23) and after forming the set of linear equations solve them for the independent coefficients in terms of the dependent one. Substitution of these values in the dependent relation gives a polynomial in the dependent coefficient which can be solved by Horner's method. Its solution then determines the independent coefficients. This procedure might be extended similarly to cases where the number of coefficients is two or more greater than that of the linear equations, but obviously the process becomes quite involved.

The values of the attenuation coefficients P_0, Q_0, etc., are unique when determined from linear equations. The impedance coefficients a_0, b_0, etc., derived from them are also single-valued to give a physical solution in most types of networks, meaning that only one such physical network has the particular attenuation characteristic. However, in the *lattice type*, it has been found that there are usually possible *two or more physical solutions* for the impedance coefficients from the attenuation coefficients, which correspond to two or more similar appearing physical structures having identically the same attenuation characteristic but different phase constants.

Method 2. Solutions from the Phase Constant

This method is applicable particularly to phase networks which ideally have no attenuation and to other networks where the number of phase coefficients equals the number of independent network parameters. The procedure is the same as in the previous method where now we operate with the phase constant formula (22). Multiplying the latter by its N-polynomial, we obtain formally the *phase linear equation*, true at all frequencies,

$$fM_1 + f^3M_3 + \cdots - IIN_0 - f^2IIN_2 - \cdots = 0. \qquad (24)$$

Fixing the phase constant, and hence II, in this equation at frequencies

equal in number to the phase coefficients gives us, if this number is equal to the number of independent network parameters, the desired set of linear equations to be solved by the usual methods. In a network where the number of phase coefficients is one less than the number of network parameters an additional relation will be needed to determine the network elements and this can be supplied from the attenuation characteristic. Here the attenuation characteristic can probably be lowered uniformly without altering the phase characteristic. (See Section 2.82.)

Method 3. Solutions from the Propagation Constant

Since it has been shown in Section 2.5 that any network of the type considered in this paper can always be represented physically by a lattice type having an equivalent propagation constant, we can simplify the discussion here by dealing entirely with the lattice network. From (13) the impedance ratio $z_{11}/2R$ for this type is derived in terms of its propagation constant as

$$\frac{z_{11}}{2R} = \frac{e^{\Gamma} - 1}{e^{\Gamma} + 1} = \tanh\,(\Gamma/2), \tag{25}$$

which holds at all frequencies. Thus, *a determination of the recurrent network from its propagation constant (attenuation and phase constants together) reduces to the solution of a two-terminal impedance network from its impedance characteristic.* The impedance ratio components s and y in (19) will become definite known functions of frequency determined through (25) by the propagation constant of the given lattice network.

A method of solving for the impedance coefficients a_0, b_0, etc., and hence the network elements from the components s and y, follows. Instead of attempting to separate the impedance ratio expression into its real and imaginary parts which can then separately be equated to s and y, which is the usual method, let us multiply (19) by the b-polynomial. Now equating separately the real and imaginary parts we obtain a pair of equations which are linear in the coefficients and hold at all frequencies. This pair of *impedance linear equations* are formally

$$\begin{aligned} &a_0 - f^2a_2 + \cdots - sb_0 + fyb_1 + f^2sb_2 + \cdots = 0, \\ \text{and} \qquad &fa_1 - f^3a_3 + \cdots - yb_0 - fsb_1 + f^2yb_2 + \cdots = 0. \end{aligned} \tag{26}$$

By this means the formulæ are put in a form such as to require in all cases the solution of a set of equations linear in the coefficients, obtained from (26) at different frequencies. A procedure for their solution

similar to that used in dealing with equation (23) can be applied and will not be repeated here. This process, apparently new, of obtaining linear equations for the impedance coefficients which contain powers of frequency and the impedance components, was applied by the writer to non-dissipative two-terminal networks in this *Journal*, January, 1923, p. 21, also in U. S. Patent No. 1,509,184, dated September 23, 1924; and to dissipative networks which simulate a smooth line impedance in U. S. Patent Application, Serial No. 134,515, filed September 9, 1926. It is merely outlined here.

2.8. *Useful Properties and Relations*

The following discussion covers a number of points concerning these networks which have been found quite useful. They can be verified readily from the fundamental formulæ and so need not be derived in detail.

2.81. *Analytical Simplifications*

Let it be desired to design a given network from its attenuation characteristic in a frequency range when the number of attenuation coefficients is one greater than the number of independent network elements. As previously stated, it is usually possible in such cases to choose as part of the attenuation data the attenuation constant at a particular frequency, such as zero or infinite frequency, and make the resulting number of attenuation coefficients and independent elements equal in number, with consequent ease of solution. Another method of simplifying the analysis might be to slightly alter the form of the given z_{11} by adding to it, or subtracting from it, a resistance element in series or in parallel. This may have the effect of making the resulting attenuation coefficients and independent elements equal in number without appreciably altering the general attenuation characteristic in the desired frequency range.

2.82. *Uniform Attenuation Change*

According to principles developed above, if the attenuation constant of a given network is changed uniformly over the entire frequency range without altering its phase constant, its distortion producing characteristics are not affected.

Let z_{11} correspond to a given *lattice type* network and z_{11}' to a derived one in which the attenuation only has been changed by a uniform amount A_0 at all frequencies. Then one form of structure for z_{11}' is

$$z_{11}' = \frac{1}{\dfrac{1}{m_1 z_{11} + m_2 R} + \dfrac{1}{m_3 R}}, \tag{27}$$

where $m_1 = \cosh^2 (A_0/2)$,

$m_2 = \sinh A_0$,

and $m_3 = 2 \coth (A_0/2)$,

m_1 being greater than unity, while m_2 and m_3 have the sign of A_0. This relation for z_{11}' stated approximately in words is as follows: *To raise the attenuation, magnify the given z_{11} and add series resistance, then add parallel resistance to the whole; to lower the attenuation, magnify z_{11} and add such negative resistances.* An example is given by Networks 1*a* and 3*a* of Appendix IV.

An impedance equivalent form of structure for z_{11}' is

$$z_{11}' = \frac{1}{\dfrac{1}{m_1' z_{11}} + \dfrac{1}{m_2' R}} + m_3' R, \tag{28}$$

where $m_1' = \operatorname{sech}^2 (A_0/2)$,

$m_2' = 4 \operatorname{cosech} A_0$,

and $m_3' = 2 \tanh (A_0/2)$,

m_1' being positive and less than unity, while m_2' and m_3' have the sign of A_0. Hence with this form, *to raise the attenuation, reduce the given z_{11} and add parallel resistance, then add series resistance to the whole; to lower the attenuation, reduce z_{11} and add such negative resistances.* An example is given by Networks 1*b* and 3*b*, Appendix IV.

It will be seen from these relations derived from a physical z_{11} that when A_0 is positive a physical z_{11}' always results. When A_0 is negative, however, physical impedances would be obtained only under certain conditions, depending upon the given z_{11} and upon A_0.

One practical utility of the relations would occur in the following situation. Suppose that a design was being attempted from assumed attenuation values with a network having such a general characteristic and that z_{11} consists of some structure in series or in parallel with a resistance element. The latter resistance as determined from the linear equations may come out to be negative and give z_{11} an unphysical structure. In such a case we could apply the above relations and raise all the attenuation values uniformly such an amount A_0 that the resulting network z_{11}' would be physical.

Corresponding relations between two networks of the *ladder type* are

$$z_{11}' = e^{A_0} z_{11} + (e^{A_0} - 1)R; \tag{29}$$

and between two of the *bridged-T* (*I*) type are

$$z_{11}' = (c \sinh (A_0/2) + \cosh (A_0/2))^2 z_{11} + 2 \sinh (A_0/2)(c \sinh (A_0/2) + \cosh (A_0/2))R,$$

and

$$c' = \frac{c + \tanh (A_0/2)}{c \tanh (A_0/2) + 1}. \tag{30}$$

In the above process we would generally be increasing the number of network parameters without changing the number or magnitude of the phase coefficients.

2.83. *Phase Constant Comparisons of Certain Pairs of Lattice Type Networks*

It has already been stated that there are usually two physical networks of the same structural lattice form which have identical attenuation constants but different phase constants. They are derivable as two physical solutions from the same attenuation coefficients. In the case of a limited class of these networks, an interesting relation exists between the phase constants of such a pair which may be stated as follows.

Theorem.—The two lattice type networks of every pair having the same attenuation characteristic in each of which the series impedance (z_{11}) *consists of a resistance in parallel with any pure reactance network, of different proportions in each, have phase constants such that their sum or difference is identical with that of a non-dissipative lattice phase network whose series impedance* (z_{11}) *is a pure reactance network proportional to that in the series impedance of either of the pair.*

A *corollary* results from this.

One network of the pair is equivalent to the tandem combination of the other and the related phase network.

It should be pointed out here that results for the case in which z_{11} is a resistance in series with a reactance network are similar, except for a phase change of π, since then the lattice impedance z_{21}, the inverse network of z_{11}, is a resistance in parallel with a reactance network.

A procedure for proving the theorem will be sketched briefly. Assume as given one network in which z_{11} is made up of a resistance in parallel with a pure reactance network whose impedance is imy, where m is a positive constant and y is a function of frequency. This gives a form

$$F = e^{2A} = \frac{1 + P_2 y^2}{1 + Q_2 y^2}. \tag{31}$$

Reversing the process, we obtain from the same coefficients P_2 and Q_2 a second similarly constructed network besides the original one. The

two physical networks differ in their phase constants but have the same attenuation constants. For one

$$\tan B' = \frac{(\sqrt{P_2} + \sqrt{Q_2})y}{-1 \quad \sqrt{P_2 Q_2 y^2}}, \tag{32}$$

and for the other

$$\tan B'' = \frac{(\sqrt{P_2} - \sqrt{Q_2})y}{1 + \sqrt{P_2 Q_2 y^2}}, \tag{33}$$

where B'' has a maximum or minimum depending upon whether y is positive or negative. As a result for the *sum*

$$\tan\left(\frac{B' + B''}{2}\right) = \sqrt{P_2}y, \tag{34}$$

and for the *difference*

$$\tan\left(\frac{B' - B''}{2}\right) = \sqrt{Q_2}y. \tag{35}$$

Now a non-dissipative lattice type network in which z_{11} is a reactance proportional to y has a formula

$$\tan (B/2) = M_1 y, \tag{36}$$

where M_1 is positive. Comparison of these latter formulæ indicates the proof of the theorem and its corollary.

A simple and useful relation exists between the maximum attenuation constant A_m occurring at $y = \infty$ and the maximum or minimum phase constant B_m'' of (33) occurring at $y = \pm 1/(P_2Q_2)^{1/4}$. It is

$$\sinh (A_m/2) = \pm \tan B_m''. \tag{37}$$

An example is given by Networks 2*a*, Appendix IV, and a practical use of this relation will be made in Section 4.2.

2.84. *Composite Networks*

The tandem combination of two or more different sections of constant resistance networks can generally give propagation characteristics which are unattainable in a single section. For this reason it is sometimes advantageous to treat such a composite network of two or three simple sections as a single unit. When this is done it will be found that the composite network has attenuation coefficients, if any, which in number may be equal to, greater than, or even less than the sum for the individual networks when considered separately.

An example of a case in which the number of attenuation coefficients

for the composite network equals the sum for the separate sections is furnished by two sections of Network 1*a* or of 2*a*, Appendix IV, both having four coefficients. On the other hand, a composite network of 1*a* and 2*a*, one of each, has five attenuation coefficients. Finally, a composite network of two sections of Network 3*a* has only five attenuation coefficients contrasted with a sum of six for the separate networks. In the latter case we can obtain only five linear equations from the attenuation characteristic which are not sufficient to determine the six series elements. This probably means that for the same attenuation characteristic the resistances in series with the two inductances can be given any ratio to each other from zero to infinity. A sixth relation can then be supplied by assuming the practical condition which makes the ratio of resistance to reactance the same in the inductance branches of both sections. This composite network can have an attenuation constant whose increase with frequency is approximately linear over a wide internal frequency range.

Composite phase networks of simple structure also lend themselves readily to such treatment as a single unit.

2.85. *Composite Lattice Networks Having Uniform Attenuation*

To a lattice type network of a certain class having a finite non-uniform attenuation characteristic there corresponds a single infinity of complementary ones, such that when any one of the latter is combined with it, the composite network has a uniform total attenuation constant and a zero total phase constant over the entire frequency range. *The separate attenuation constants are complementary while the phase constants are equal, but opposite in sign.* Such a conposite network we have seen would be *absolutely distortionless.* It is a relatively simple matter to obtain the necessary relations which such a complementary network must bear to the first if we impose these propagation conditions on the combination. Two sets of relations may be derived, each corresponding to a particular structure for the first network, with the following results.

If the given section (A, B) has *series impedances*

$$z_{11} = R_s + z_s, \tag{38}$$

where R_s is a resistance and z_s is any impedance, any equivalent transformation of which does not contain series resistance, and if a complementary network (A', B') is added such as to give a composite network (A_c, B_c) with the propagation constant

$$A_c = A + A' = \text{constant},$$

and (39)

$$B_c = B + B' = 0,$$

then the complementary network is given by

$$z_{11}' = R_1 + \frac{1}{\frac{1}{z_2} + \frac{1}{R_3}}, \tag{40}$$

where $R_1 = 2 \coth (A_c/2)R$,

$z_2 = 4 \operatorname{cosech}^2 (A_c/2)R^2/z_s$,

and $R_3 = 4 \operatorname{cosech}^2 (A_c/2)R^2/(R_s - 2 \coth (A_c/2)R)$.

Here z_2 is the inverse network of z_s of impedance product $4\operatorname{cosech}^2 (A_c/2)R^2$. The network in (40) is R_1 in series with the parallel combination of z_2 and R_3. An equivalent form for z_{11}' is

$$z_{11}' = \frac{1}{\frac{1}{R_1' + z_2'} + \frac{1}{R_3'}}, \tag{41}$$

where $R_1' = \cosh^2 (A_c/2)(R_s - 2 \tanh (A_c/2)R)$,

$z_2' = \cosh^2 (A_c/2)(R_s - 2 \tanh (A_c/2)R)^2/z_s$,

and $R_3' = \dfrac{2R(\coth (A_c/2)R_s - 2R)}{(R_s - 2 \coth (A_c/2)R)}$.

It will be a *physical network* provided A_c satisfies the relation

$$1 < \coth (A_c/2) \leqq R_s/2R. \tag{42}$$

At the *minimum* A_c, $R_1 = R_1'$, $z_2 = z_2'$, and $R_3 = R_3' = \infty$.

If, on the other hand, the given section has *parallel impedances* (similar to the preceding network of (38) whose output terminals are reversed),

$$z_{11} = \frac{1}{\frac{1}{R_p} + \frac{1}{z_p}}, \tag{43}$$

where R_p is a resistance and z_p is any impedance, any equivalent transformation of which does not contain parallel resistance, then a corresponding complementary network has one form given by

$$z_{11}' = \frac{1}{\frac{1}{R_1} + \frac{1}{z_2 + R_3}}, \tag{44}$$

where $R_1 = 2 \tanh (A_c/2)R$,

$z_2 = 4 \sinh^2 (A_c/2)R^2/z_p$,

and $R_3 = 2 \sinh^2 (A_c/2)R(2R - \coth (A_c/2)R_p)/R_p$.

An equivalent form is

$$z_{11}' = \frac{1}{\frac{1}{R_1'} + \frac{1}{z_2'}} + R_3', \tag{45}$$

where $R_1' = 2 \operatorname{sech}^2 (A_c/2)RR_p/(2R - \tanh (A_c/2)R_p)$,

$z_2' = 4 \operatorname{sech}^2 (A_c/2)R^2R_p^2/(2R - \tanh (A_c/2)R_p)^2 z_p$,

and $R_3' = \dfrac{2R(2R - \coth (A_c/2)R_p)}{(2 \coth (A_c/2)R - R_p)}$.

There will be a *physical network* provided

$$1 < \coth (A_c/2) \leqq 2R/R_p. \tag{46}$$

At the *minimum* A_c, $R_1 = R_1'$, $z_2 = z_2'$, and $R_3 = R_3' = 0$.

It may be added that if (38) and (43) represent inverse networks of impedance product $4R^2$, then another such pair is given by (40) and (44), and still another by (41) and (45).

An extension of these results may now readily be made to give *two-section composite networks whose attenuation constants are uniform but whose phase constants are not zero.* It has been stated that to every lattice type network having finite attenuation there usually corresponds another one of the same structural form having the same attenuation but a different phase characteristic. Hence, in either case above where the two complementary sections giving a total uniform attenuation are known, we may derive by regular methods the alternative lattice sections, having, respectively, the same attenuation constants. Since we would then have two sections to give the one attenuation characteristic and two sections for the complementary characteristic, it would be possible to obtain *four composite networks* of similar structure, all of which give *the same uniform attenuation but four different phase characteristics.* One of these combinations would be the case in which the phase constant is zero. Four more phase characteristics, differing from the others by an amount π, can obviously be obtained by reversing the terminals of either section.

2.9. *Procedure for the Design of Distortion Correcting Networks*

It would be most gratifying to be able to obtain directly from a desired propagation characteristic the corresponding form of network.

This is generally a difficult problem and it becomes necessary to resort to simplifying methods somewhat similar to those employed in the design of electric wave-filters. One reason for this difficulty is that we are limited to physical resistance, inductance, and capacity elements, all of which must, in general, be positive. We would, therefore, begin with known forms of networks whose general propagation characteristics have been determined and choose from them one or more whose combination offers the possibility of giving a satisfactory desired result. A number of points which are applicable in the general case may be noted as follows:

1. First, determine the desired propagation characteristics of the distortion correcting network corresponding to formula (8).

2. If necessary, divide this propagation characteristic into several parts each of which has the approximate characteristic belonging to a known network structure.

3. Assume one of these networks physically capable of having such an alloted characteristic and attempt a design to approximately fit it according to one of the methods of Section 2.7. Where there is an attenuation characteristic, Method 1 is usually best, as attenuation is generally of more importance than phase and hence its simulation requires greater accuracy. The network will introduce a phase constant which will necessarily have to be taken into account. Of the two or more possible solutions for the lattice type network, the one with the most desirable phase constant would obviously be chosen and in some cases this may be close to requirements. Another reason for usually following this order of simulating the attenuation first and the resultant phase later is furnished as a consequence of Theorems I and II of Section 2.2. From them we see the physical possibility of introducing certain phase characteristics without attenuation (ideally), but not varying attenuation characteristics without phase. Method 3 imposes a rather severe requirement on a single network.

4. If the network design comes out to be unphysical with the particular characteristic values assumed, small variations from these values should be tried, since the natural varying curvatures in the propagation characteristic of the network must sometimes be allowed for. Otherwise, a different kind of network should be used, or a composite one, which has a similar characteristic.

5. In designing successive sections of the complete transducer, the effects of previous parts must be considered.

To facilitate the application of this method of distortion correction, general propagation characteristics together with formulæ have been derived for a representative number of lattice type structures. These

are given in Appendix IV. Any pair of the networks, such as 1*a* and 1*b*, differ only by an interchange of series and lattice elements with a corresponding difference in their phase constants of an amount π. In order to simplify computations for some networks the formulæ were derived so as to require attenuation data at a limiting frequency, but other formulæ may also be obtained. By means of the relations in Section 2.5, transformations can readily be made to any of the other general types, if they lead to physical structures.

The type of network which a final design is to assume will be suggested by economic and practical considerations. However, an approximate statement can be made in this connection. If the sections are to be dissymmetrical as regards the two pairs of terminals and unbalanced as regards the two sides of the line, use the full-series or full-shunt ladder types; if symmetrical and unbalanced, use the bridged-T types; if symmetrical and balanced, use the bridged-T or lattice types.

Part 3. Arbitrary Impedance Recurrent Networks

In Part 2 consideration was given entirely to recurrent networks whose iterative impedances are a constant resistance at all frequencies and which depend upon the use of inverse networks; that is, $z_{11}z_{21} = R^2$. It is intended here merely to point out briefly that all the types in Section 2.4 can be generalized to have iterative impedances of arbitrary value K provided in them

$$R \text{ is generalized to } K,$$

and

$$z_{11}z_{21} = K^2; \tag{47}$$

that is, z_{11} and z_{21} are *inverse networks* [12] *of impedance product* K^2. The corresponding propagation constant formulæ hold also with these generalizations.

Where a recurrent network of arbitrary iterative impedance K is desirable, these structures would, theoretically at least, be applicable. Practically, however, considerable difficulties are usually encountered in physically realizing z_{11} and z_{21} to give a desired propagation constant, and perhaps even K when K is not a simple function of frequency. A few physical possibilities will be given here in which the structures for z_{11} and z_{21} are easily identified from the forms of the expressions. They may be used in the different types of networks, and, of course, z_{11} and z_{21} may be interchanged.

[12] The complete qualifying statement such as given is necessary here, not just simply "inverse networks."

1.

$$K = R + iL\omega;$$
$$z_{11} = iL_{11}\omega, \tag{48}$$

and

$$z_{21} = (2RL/L_{11}) + i(L^2/L_{11})\omega + 1/i(L_{11}/R^2)\omega.$$

The impedance z_{21} is series resistance, inductance and capacity.

2.

$$K = R + 1/iC\omega;$$
$$z_{11} = 1/iC_{11}\omega, \tag{49}$$

and

$$z_{21} = (2RC_{11}/C) + i(R^2C_{11})\omega + 1/i(C^2/C_{11})\omega.$$

Here z_{21} is the same type of structure as in (48).

3.

$$K = R + 1/iC\omega;$$
$$z_{11} = \frac{1}{\dfrac{1}{\dfrac{R}{m_1} + \dfrac{1}{im_1C\omega}} + \dfrac{1}{\dfrac{R}{m_2} + \dfrac{1}{im_2C_{11}\omega}}}, \tag{50}$$

and

$$z_{21} = \frac{1}{\dfrac{1}{R_{21}} + \dfrac{1}{iL_{22}\omega}} + R_{23} + \frac{1}{iC_{24}\omega},$$

where $R_{21} = m_2R(C - C_{11})^2/C^2$,

$$L_{22} = m_2R^2C_{11}(C - C_{11})^2/C^2,$$
$$R_{23} = R(m_1C^2 + m_2C_{11}(2C - C_{11}))/C^2,$$

and

$$C_{24} = C^2/(m_1C + m_2C_{11}).$$

The impedance z_{11} consists of two parallel branches each containing series resistance and capacity; z_{21} is made up of parallel resistance and inductance in series with both resistance and capacity. For certain values of the parameters m_1, m_2, and C_{11}, even though positive, the resistance R_{23} can become negative and hence unphysical as a passive element.

A lattice or equivalent network made up of such impedances, in addition to having the assumed iterative impedance which approximates that of an open-wire line at the upper frequencies, can have an attenuation constant decreasing with frequency which tends to equalize that of a length of such line; the attenuation formula has the form

$$F = e^{2A} = \frac{P_0 + P_2f^2}{Q_0 + f^2}. \tag{51}$$

4.

$$K = \sqrt{\frac{R' + iL'\omega}{G' + iC'\omega}}, \quad \text{(smooth line)};$$

$$z_{11} = mR' + imL'\omega, \tag{52}$$

and

$$z_{21} = \frac{1}{(mG' + imC'\omega)}.$$

Another possible simple pair is that in which z_{11} is a resistance and z_{21} is either series resistance and inductance in parallel with series resistance and capacity or parallel resistance and inductance in series with parallel resistance and capacity. These impedance elements may be used in the lattice or bridged-T (II) type structures where the impedance element K is not explicitly required. Extension to more complex structures can be made by the methods of Section 2.3. An application will be given in Section 4.7 which considers the simulation of a smooth line.

Owing to the much greater inherent difficulty of physically realizing inverse networks of impedance product K^2 when K is not R, the generalization does not add much practically for our purpose, but some structures in which K is not R may be of utility under particular conditions.

Part 4. Applications

4.1. *Complementary Distortion Correcting Networks*

The pair of networks in Fig. 8 illustrates in a very simple manner the general relations given in Section 2.85, as well as ideal distortion

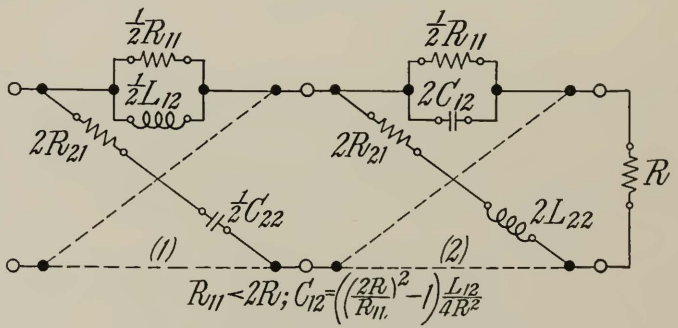


Fig. 8—Distortionless composite network.
(Broken lines indicate the other series and lattice branches, respectively identical).

correction over the entire frequency range. When placed in tandem they represent a composite network whose attenuation constant is uniform at all frequencies and whose phase constant is zero, which are

characteristics for no distortion. Let us obtain the steady-state characteristics of each network and of the composite one; then consider transient conditions and obtain the indicial voltages of the corresponding networks to verify again by this illuminating example that the steady-state characteristics laid down for no distortion are quite sufficient when transient conditions exist.

The first section is Network 2*a*, Appendix IV, wherein z_{11} is parallel resistance R_{11} and inductance L_{12}, with R_{11} less than $2R$ and the characteristic 1. Let us put $m = R_{11}/2R$, and $n = L_{12}/2R$. Then

$$\frac{z_{11}}{2R} = \frac{imn\omega}{m + in\omega}, \tag{53}$$

and the propagation constant formula becomes from (13)

$$e^{\Gamma_1} = \frac{m + i(1 + m)n\omega}{m + i(1 - m)n\omega}. \tag{54}$$

To obtain a complementary second section let us assume that the total attenuation constant, A_c napiers, of the composite structure is to equal the maximum of the first section which occurs at infinite frequency. Then from the above

$$e^{A_c} = \frac{1 + m}{1 - m}$$

and $\tanh (A_c/2) = m$, giving as the correcting section by (44) one of Network 1*b*, Appendix IV, with characteristic 1 in which

$$R_{11} = 2mR, \text{ as in (53)},$$

and

$$C_{12} = \frac{(1 - m^2)n}{2m^2R} = \left[\left(\frac{2R}{R_{11}}\right)^2 - 1\right]\frac{L_{12}}{4R^2}. \tag{55}$$

For this second section then

$$\frac{z_{11}}{2R} = \frac{m^2}{m + i(1 - m^2)n\omega},$$

and

$$e^{\Gamma_2} = \left(\frac{1 + m}{1 - m}\right)\left(\frac{m + i(1 - m)n\omega}{m + i(1 + m)n\omega}\right). \tag{56}$$

Obviously, from (54) and (56),

$$e^{\Gamma_1 + \Gamma_2} = \frac{1 + m}{1 - m} = e^{A_c},$$

as was assumed.

The attenuation and phase constants of each of these two sections and the combined structure are shown in Fig. 9, as a function of $L_{12}\omega/2R$, where $R_{11} = R$. It will be seen that A_1 and A_2 are complementary while B_1 and B_2 are equal but opposite. For the composite network A_c = constant and $B_c = 0$; thus the latter phase constant

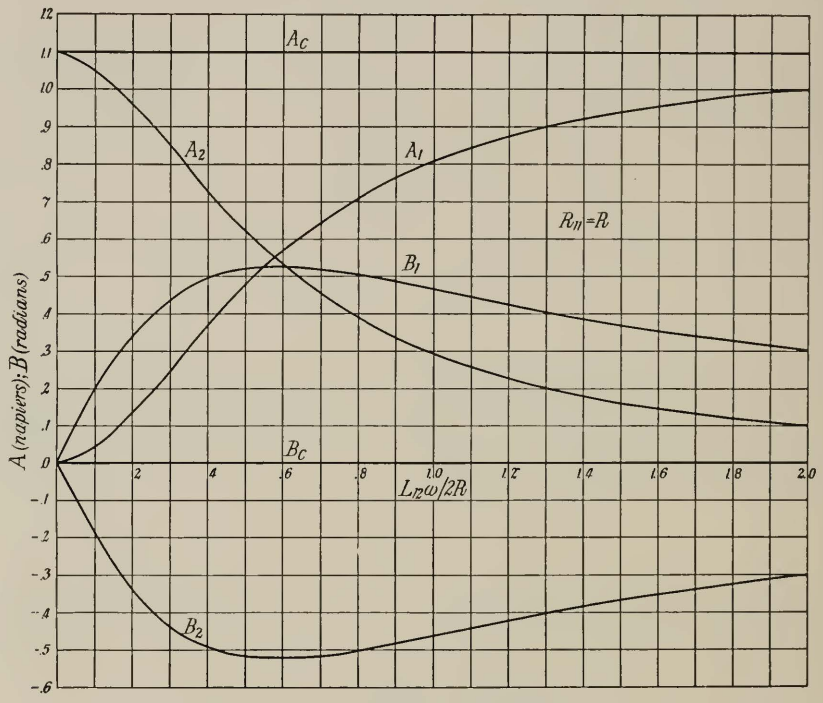


Fig. 9—Propagation constants in distortionless network.

has a zero slope with frequency. Whatever steady periodic voltage exists at one end would appear across the terminating resistance R in the same phase but attenuated by an amount A_c napiers. Since these conditions hold for the composite network at all frequencies, we should expect to obtain for it an indicial voltage and time-of-transmission, respectively,

$$g_c(t) = e^{-A_c},$$

and (57)

$$\tau = \frac{B_c}{\omega} = \frac{dB_c}{d\omega} = 0.$$

Let us next determine the indicial voltages of the individual sections when each is closed by a resistance R. Substitute the operator p for $i\omega$ and obtain symbolically from (54) and (56)

$$e^{-\Gamma_1} = 1 - 2mn\left(\frac{p}{m + (1 + m)np}\right), \tag{58}$$

and

$$e^{-\Gamma_2} = \frac{1 - m}{1 + m} + \frac{2mn(1 - m)}{1 + m}\left(\frac{p}{m + (1 - m)np}\right). \tag{59}$$

Introducing these expressions in the general relation, where the network is terminated by R,

$$\frac{e^{-\Gamma}}{p} = \int_0^\infty e^{-pt} g(t)dt, \tag{60}$$

there results for the indicial voltage of the first section, since

$$\frac{1}{u + vp} = \int_0^\infty e^{-pt}\left(\frac{e^{-(ut/v)}}{v}\right)dt, \tag{61}$$

$$g_1(t) = 1 - \frac{2m}{1 + m}e^{-[mt/(1+m)n]}; \tag{62}$$

and for the second section

$$g_2(t) = \frac{1 - m}{1 + m} + \frac{2m}{1 + m}e^{-[mt/(1-m)n]}. \tag{63}$$

These functions are given in Fig. 10.

It will now be shown that, whereas the indicial voltage of each section alone is a varying function of time, that of the composite network is a constant, which represents the transient condition for no distortion with zero time-of-transmission.

For the composite network terminated by R the indicial voltage $g_c(t)$ may be derived from the usual formula for such a combination, equivalent to (5),

$$g_c(t) = g_2(0)g_1(t) + \int_0^t g_1(t - y)g_2'(y)dy. \tag{64}$$

Upon carrying through the integration we get

$$g_c(t) = \frac{1 - m}{1 + m} = e^{-A_c} = \text{constant}, \tag{65}$$

which agrees with the prediction from the steady state and is so shown in Fig. 10.

Obviously the two sections can be interchanged.

The composite network appears at first hand to behave in a rather remarkable manner. For if a periodic voltage is suddenly impressed

at one end, the steady state will not be established within the network until after some lapse of time, whereas it occurs at the terminating resistance instantaneously. This property is, of course, to be expected from its steady-state characteristics.

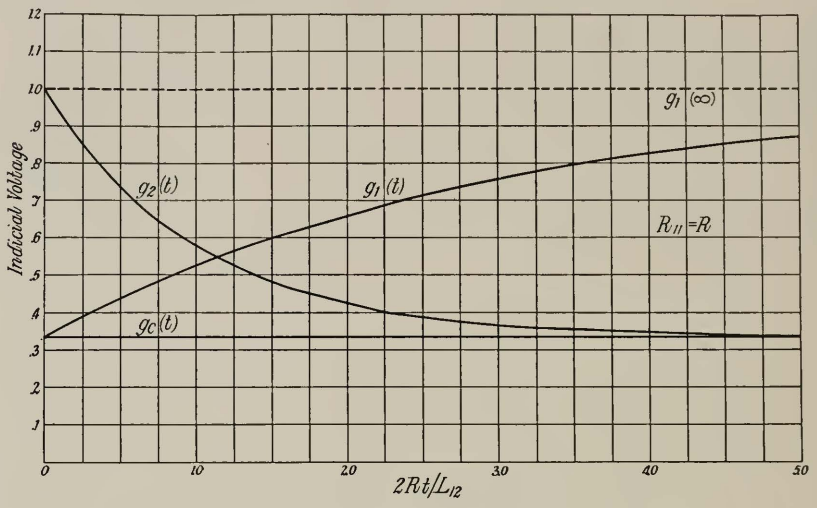


Fig. 10—Indicial voltages in distortionless network.

It may be added that such networks would still give complementary results if separated for any purpose by a symmetrical line in a circuit which is terminated at each end by a resistance R and which has an e.m.f. applied through one of the resistances. The separation of the two complementary networks under these conditions would result in the same current being received by the terminating resistance as when both networks are together at one end, where it is known the networks would produce no distortion. This follows immediately from the reciprocal theorem. For by it we readily see that the same current would be transmitted to the input terminals of the complementary receiving network whether the first network was at one end or the other. (These two cases are equivalent from the standpoint of received current to turning the combined transmission line and first network end for end.)

4.2. *Distortion Correction in Submarine Cable Circuit*

The following illustration shows the improvement which can be made in the shape of the arrival voltage at the end of a long submarine cable circuit by distortion correction at the very low frequencies only. Such an improvement would increase the speed of building up of d-c. telegraph signals and hence allow a greater speed of signaling.

The circuit assumed is a submarine cable whose length, l, is 1700 miles and whose parameters are to have the constant values per mile

$$R' = 2.74 \text{ ohms}; \qquad L' = .001 \text{ h.};$$
$$G' = 0 \qquad ; \qquad C' = .296 \text{ mf.}$$

It is terminated at the receiving end only by a resistance $R = \sqrt{L'/C'} = 58.12$ ohms. The transfer exponent, $a + ib$, of this circuit at the terminal resistance is computed from the formula, easily derived,

$$e^{a+ib} = (k/R) \sinh \gamma l + \cosh \gamma l, \tag{66}$$

where

$$\gamma = \sqrt{(R' + iL'\omega)iC'\omega},$$

and

$$k = \sqrt{(R' + iL'\omega)/iC'\omega}.$$

These results are shown in Fig. 12.

It is desired to obtain distortion correction in this circuit from 0 to 25 cycles per second by introducing a terminal constant resistance transducer which will approximately equalize the attenuation over this range and make the resultant phase linear with frequency. Since in practice there is interference between different cables at higher frequencies, the correcting network should introduce increased attenuation above this range. Calculations gave

$$\text{at } f = 0, \; a = 4.40 \text{ napiers};$$

and

$$\text{at } f = 25\sim, \; a = 14.10 \text{ napiers.}$$

Assuming arbitrarily that the network will have at $f = 25\sim$ an attenuation of only .30 napier, the ideal total attenuation for the frequency range is

$$a' = 14.10 + .30 = 14.40 \text{ napiers.} \tag{67}$$

The attenuation of the network should decrease from a maximum value of $(14.40 - 4.40) = 10.00$ napiers at $f = 0$ to a value of .30 napier at $f = 25\sim$ and then increase with frequency. If a linear relation for the resultant phase is assumed so as to cross the b curve at about $f = 25\sim$, the phase which the network should give is negative in the range with a minimum of about -2.75 radians, and is zero at $f = 0$ and $f = 25\sim$.

A network having this desired general type of propagation constant is Network 8, Appendix IV, with the characteristic 1, but a single section will not be sufficient since its minimum phase is between 0

and $-\pi/2$ radians. The best number of sections to use is determined by the total minimum phase required and can be found here quite readily, as follows. Because of the comparatively small amount of attenuation assumed for the total correcting network at $f = 25\sim$, this type of network is one in which z_{11} consists of a resistance in parallel with an approximate reactance so that we may apply for the present purpose the relation (37) between maximum attenuation and minimum phase of such a section. For a total maximum attenuation of 10.00 napiers this relation gives for two sections a total minimum phase of $-$ 2.81 radians, which is close to the required value $-$ 2.75 radians. Three sections give $-$ 3.59 radians, showing the best number to be two. (If the result with two identical sections had been a negative phase considerably greater than the required value, it would have been possible to proportion the total maximum attenuation at zero frequency between two such different sections so as to give approximately the desired total minimum phase. In such a case each section could be designed from its corresponding proportion of the total attenuations at the other frequencies.)

Each of two such identical sections was designed by the formulæ given in Appendix IV, using attenuation data fixed by the values of $(a' - a)/2$. Allowances had to be made at $f_1 = 5\sim$ and $f_2 = 15\sim$ for necessary curvature in the attenuation characteristic so as to obtain a physical result. It was assumed that the phase constant would turn out to be satisfactory since it had already been given some consideration when determining the number of sections. The frequencies and corresponding attenuations used were

$$\begin{aligned} f_0 &= 0, & A_0 &= 5.00 \text{ napiers;} \\ f_1 &= 5\sim, & A_1 &= 3.25 \text{ napiers;} \\ f_2 &= 15\sim, & A_2 &= 1.78 \text{ napiers;} \\ f_3 &= 25\sim, & A_3 &= .15 \text{ napier.} \end{aligned}$$

The solution of the attenuation linear equations gave

$$P_2 = -68.737; \qquad Q_2 = 1.1929; \qquad Q_4 = 2.5537 \cdot 10^{-6}.$$

Whence

$$a_0 = .98661 \; ; \qquad a_1 = 1.1854 \cdot 10^{-3};$$

$$b_1 = 15.829 \cdot 10^{-3}; \qquad b_2 = 1.5980 \cdot 10^{-3}.$$

Also,

$$R_{11} = 9.42 \text{ ohms}; \qquad L_{12} = 1.994 \text{ h.};$$

$$C_{13} = 20.30 \text{ mf.}; \qquad R_{14} = 114.68 \text{ ohms};$$

where $R = 58.12$ ohms.

These results were transformed to give a ladder type network according to Section 2.5 and then incorporated in two of the dissymmetrical unbalanced full-shunt sections, as shown in Fig. 11.

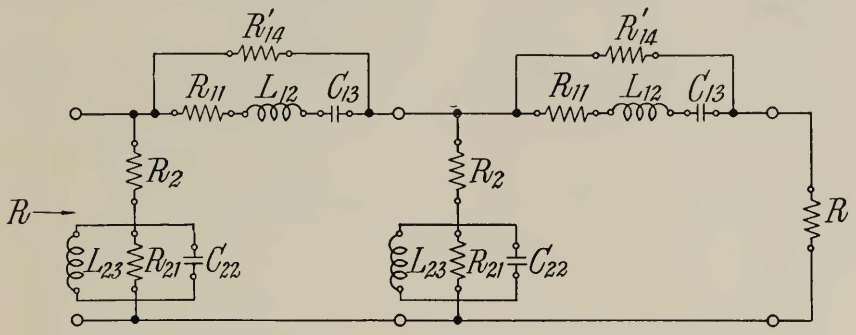


Fig. 11—Distortion correcting network for submarine cable circuit.

This transformation gives a different parallel resistance in the series branch, namely,

$$R_{14}' = 2a_0R/(1 - a_0). \tag{68}$$

Here $R_{14}' = 8565$ ohms. The elements of z_{21} in the shunt branch of the ladder type were determined from the inverse network relations

$$R_{11}R_{21} = L_{12}/C_{22} = L_{23}/C_{13} = R_{14}'R_{24}' = R^2.$$

Finally combining two resistances which are in series, $R_2 = R + R_{24}'$, we have

$$R_{21} = 359 \text{ ohms}; \qquad C_{22} = 590.3 \text{ mf.};$$

$$L_{23} = .0686 \text{ h.}; \qquad R_{24}' = .39 \text{ ohm};$$

and $$R_2 = 58.51 \text{ ohms}.$$

In Fig. 12 are shown the steady-state propagation characteristics of the uncorrected circuit, the correcting network, and the corrected circuit; the latter indicates approximately ideal conditions up to 25 cycles per second.

The improvement in shape of the arrival voltage due to this distortion correction can be seen from Fig. 13 which gives the ratio of indicial to final voltage for both the uncorrected and corrected circuit, a constant e.m.f. being impressed at the sending end at time $t = 0$. (These were computed from the steady-state characteristics of the respective circuits, using formulæ based upon those given by J. R. Carson in *B. S. T. J.*, 1924, p. 563.) The building-up speed has been increased, perhaps fourfold. The arrival voltage for the corrected circuit is

within 3 per cent of its final value when that for the uncorrected circuit has reached but half value. The initial maximum in the former is similar to that in the case of a low-pass wave-filter [13] and may be

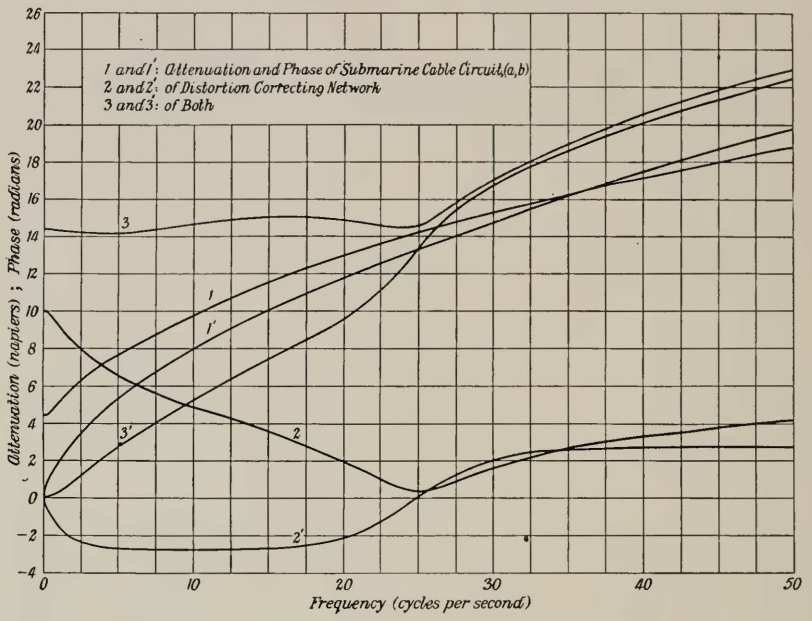


Fig. 12—Transmission characteristics of submarine cable circuit and distortion correcting network.

due to the increasing attenuation beyond the equalized range. It is probable that had but partial equalization been obtained without a

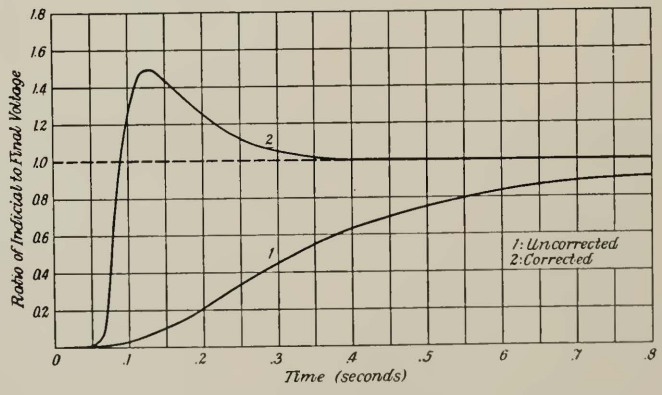


Fig. 13—Ratio of indicial to final voltage for (1) uncorrected and (2) corrected submarine cable circuit.

[13] "Transient Oscillations in Electric Wave-Filters," J. R. Carson and O. J. Zobel, *B. S. T. J.*, July, 1923.

sharp change in the attenuation, such a maximum would not have been produced. However, it is desirable to sharply attenuate the higher frequencies as has been done here, for the reason stated above. It is of interest to point out that the time-of-transmission which might be expected for the corrected circuit from the low-frequency slope with angular frequency of the steady-state phase, approximately $\tau = .076$ second, is actually the time at which the indicial voltage increases most rapidly and has reached about .4 its final value, a quite satisfactory agreement.

4.3. *Distortion Correction in Loaded-Cable Program Transmission Circuits*

Circuits which transmit programs originating at distant points to a radio broadcasting station need to be of considerably better quality over a wider frequency range than those used for ordinary telephone transmission and must be reliable under various weather conditions. Such circuits can be obtained economically with lightly loaded cable pairs which have been corrected by terminal networks for each repeater section.

The design of distortion correcting networks applicable to a 50-mile repeater section of 16-gauge H-44 cable follows. The section is terminated at *each end* by a resistance $R = 600$ ohms, the generator which impresses the voltage E having an internal impedance R. Since the received voltage would be only $.5E$ with the cable removed, in this case we are interested in the ratio

$$\frac{v}{.5E} = e^{-a-ib},$$

where a then represents the *insertion loss* in napiers.

If Γ and K are the propagation constant and iterative impedance (here used at mid-section) of the loaded cable [14] of length, l, it can be shown that the transfer exponent is

$$a + ib = \Gamma l + M_1 + M_2, \tag{69}$$

where Γl = propagation length,

$$e^{M_1} = \frac{1}{2}\left[1 + \frac{1}{2}\left(\frac{K}{R} + \frac{R}{K}\right)\right],$$

and

$$e^{M_2} = 1 - \left(\frac{K - R}{K + R} e^{-\Gamma l}\right)^2.$$

The above, of course, includes the effects of circuit terminations.

[14] Accurate computations for the propagation constant of the loaded cable were made readily by means of an improved formula for $\cosh^{-1}(x + iy)$, given in Appendix III.

It was desired to equalize the attenuation over a frequency range from zero to 4500 cycles per second and improve the time-of-phase-transmission at the lower frequencies. Computations for this 50-mile cable circuit gave values of attenuation (a in T.U.) and time-of-phase-transmission ($b/2\pi f$) as shown in Fig. 15. These circuit characteristics suggested the use of two different networks in tandem shown separately in Fig. 14, one equalizing principally at the lower frequencies, the other at the higher frequencies of the required range.

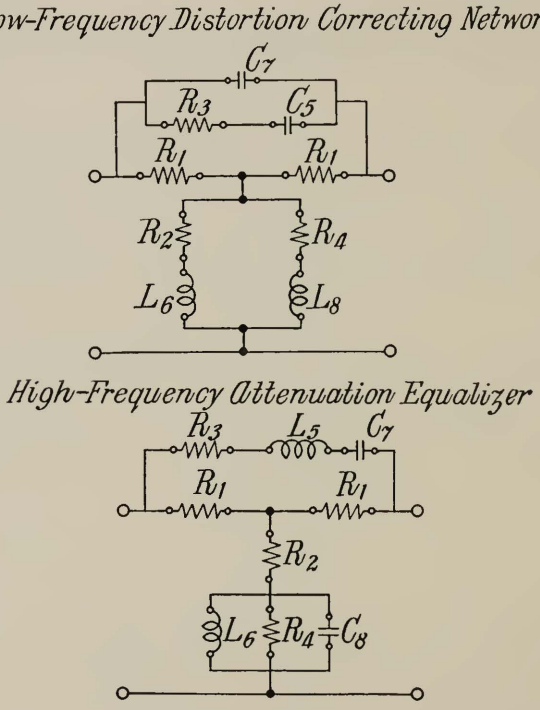


Fig. 14—Distortion correcting networks for program transmission circuit.

The low-frequency correcting network, shown as the upper section in Fig. 14, is of the symmetrical unbalanced bridged-T (Ia) type and was transformed from Network 7, Appendix IV. In the design of the latter the attenuation data corresponding to (8) were

$$f_1 = 40\sim, \quad A_1 = .536 \text{ napier};$$
$$f_2 = 200\sim, \quad A_2 = .291 \text{ napier};$$
$$f_3 = 800\sim, \quad A_3 = .176 \text{ napier};$$
$$f_4 = 2000\sim, \quad A_4 = .100 \text{ napier}.$$

Solution of the resulting four attenuation linear equations gave

$$P_0 = 102.007 \cdot 10^9; \qquad P_2 = 5.06037 \cdot 10^6;$$
$$Q_0 = 32.200 \cdot 10^9; \qquad Q_2 = 3.43087 \cdot 10^6;$$

from which

$$a_0 = .28054; \qquad a_1 = .88319 \cdot 10^{-3};$$
$$b_1 = 8.6884 \cdot 10^{-3}; \qquad b_2 = 4.0094 \cdot 10^{-6}.$$

Then, where $R = 600$ ohms, the series elements in the lattice structures are

$$R_{11} = 248.40 \text{ ohms}; \qquad C_{12} = 2.0171 \text{ mf.};$$
$$C_{13} = .6021 \text{ mf.}; \qquad R_{14} = 336.65 \text{ ohms}.$$

Transforming from this lattice type to the equivalent bridged-T (I*a*) type, we eliminate a parallel resistance in the bridged series branch (corresponding to R_{14}) by letting

$$c = 1/a_0. \tag{70}$$

Then in Fig. 14, where $c = 3.5645$,

$$R_1 = 168.3 \text{ ohms}; \qquad R_3 = 248.4 \text{ ohms};$$
$$C_5 = 2.0171 \text{ mf.}; \qquad C_7 = .6021 \text{ mf.};$$

and in the shunt branch

$$R_2 = 3037.4 \text{ ohms}; \qquad R_4 = 1458.1 \text{ ohms};$$
$$L_6 = .243 \text{ h.}; \qquad L_8 = 2.010 \text{ h.}$$

This latter useful form in which resistances are in series with inductances was obtained from the regular bridged-T (I*a*) shunt elements by means of Transformation *C*, *B. S. T. J.*, January, 1923, p. 45.

The high-frequency network, shown as the lower section in Fig. 14, is well suited to extend the range of attenuation equalization above that so far considered and was derived from Network 8, Appendix IV. Allowing for both cable and low-frequency network attenuations, and arbitrarily assuming this network to have an attenuation of .300 napier at 4500 cycles per second, the data became (as from (8))

$$f_0 = 0, \qquad A_0 = .796 \text{ napier};$$
$$f_1 = 3000\sim, \qquad A_1 = .747 \text{ napier};$$
$$f_2 = 4000\sim, \qquad A_2 = .530 \text{ napier};$$
$$f_3 = 4500\sim, \qquad A_3 = .300 \text{ napier}.$$

The solution is

$$P_2 = -46.207 \cdot 10^{-8}; \qquad Q_2 = -9.0092 \cdot 10^{-8}; \qquad Q_4 = 23.198 \cdot 10^{-16}.$$

Whence

$$a_0 = .37824; \qquad a_1 = 8.4245 \cdot 10^{-6};$$
$$b_1 = 57.522 \cdot 10^{-6}; \qquad b_2 = 4.8164 \cdot 10^{-8}.$$

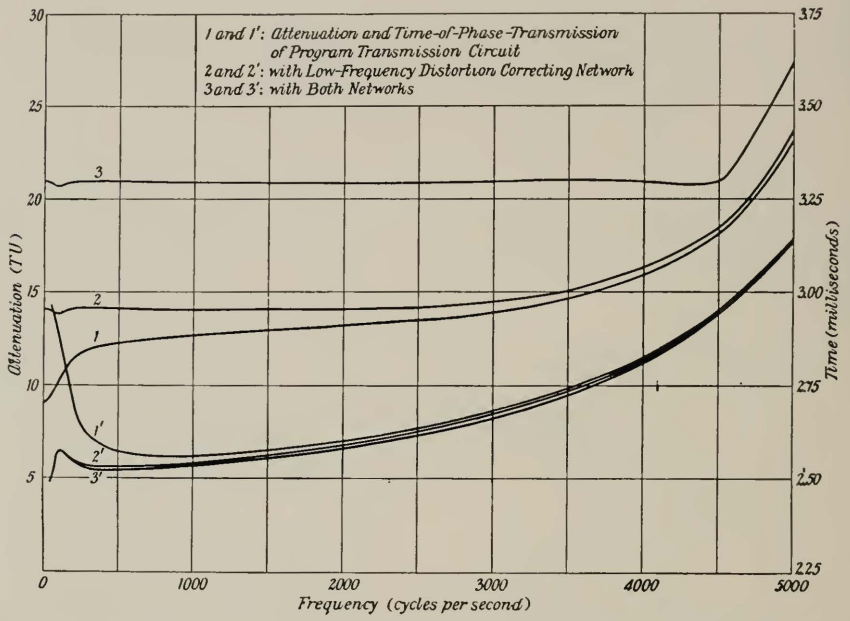


Fig. 15—Transmission characteristics of program transmission circuit with and without distortion correcting networks.

The series elements of the lattice structure are

$$R_{11} = 286.8 \text{ ohms}; \qquad L_{12} = .0987 \text{ h.};$$
$$C_{13} = .01236 \text{ mf.}; \qquad R_{14} = 453.9 \text{ ohms}.$$

Transforming to the equivalent bridged-T (Ia) type, we take c similarly as in (70); thus $c = 2.6438$. The series elements in Fig. 14 then become

$$R_1 = 226.9 \text{ ohms}; \qquad R_3 = 286.8 \text{ ohms};$$
$$L_5 = .0987 \text{ h.}; \qquad C_7 = .01236 \text{ mf.};$$

and the shunt elements

$$R_2 = 679.7 \text{ ohms}; \qquad R_4 = 1255.0 \text{ ohms};$$
$$L_6 = .00445 \text{ h.}; \qquad C_8 = .2741 \text{ mf.}$$

The effect of adding these two sections successively to the cable circuit is shown in Fig. 15. It will be seen that the first section, besides equalizing the attenuation up to about 2000 cycles per second, produces as well approximately ideal results on the time-of-phase-transmission at the lower frequencies. The complete circuit attenuation departs less than .2 T.U. from a constant value everywhere over the assumed frequency range. If desired, the time-of-phase-transmission could be improved also at the upper frequencies by the addition of proper phase networks. Such a type of correction will be made in the following application.

4.4. *Distortion Correction in Open-Wire Television Circuit*

The networks to be described here were designed by the writer especially for the particular open-wire circuit which was used for the television demonstrations from Washington, D. C., to New York City on April 7, 1927. They were designed entirely from calculated data, some of which had previously been derived from measurements on other similar lines, as the complete circuit was not available for measurements until later.

The circuit had a total length of about 285 miles, being made up principally of 276.4 miles of 165-mil open-wire pair together with 8.43 miles of necessary entrance, submarine and underground 13-gauge carrier-loaded cable (C-4.1). The iterative impedances of these two types of lines are very nearly the same in the frequency range considered and were so assumed in what follows. Hence, the propagation length of the circuit was taken as the sum of the propagation lengths of the parts. In order that such a circuit be suitable for television transmission it must be made to have extremely high quality over a very wide frequency range by means of distortion correcting networks. The requirements which the design of the present networks aimed to meet follow.

Design Requirements

1. An impedance of 600 ohms is to terminate the line at each end.

2. The attenuation, or insertion loss, of the corrected circuit is to be constant within $\pm$ 1 T.U. over the entire frequency range from 10 to 20,000 cycles per second.

3. The time-of-phase-transmission of the corrected circuit is to be constant within $\pm$ 500 microseconds (10^{-6}) from 10 to 400 cycles per second, and to be the same constant within $\pm$ 10 microseconds from 400 to 20,000 cycles per second.

4. Provision is to be made for distortion correction under various

weather conditions of the open-wire line. Details of the process of arriving at some of these requirements, also measurements and performance of the complete television circuit, have been given in a previous number of this *Journal*.[15]

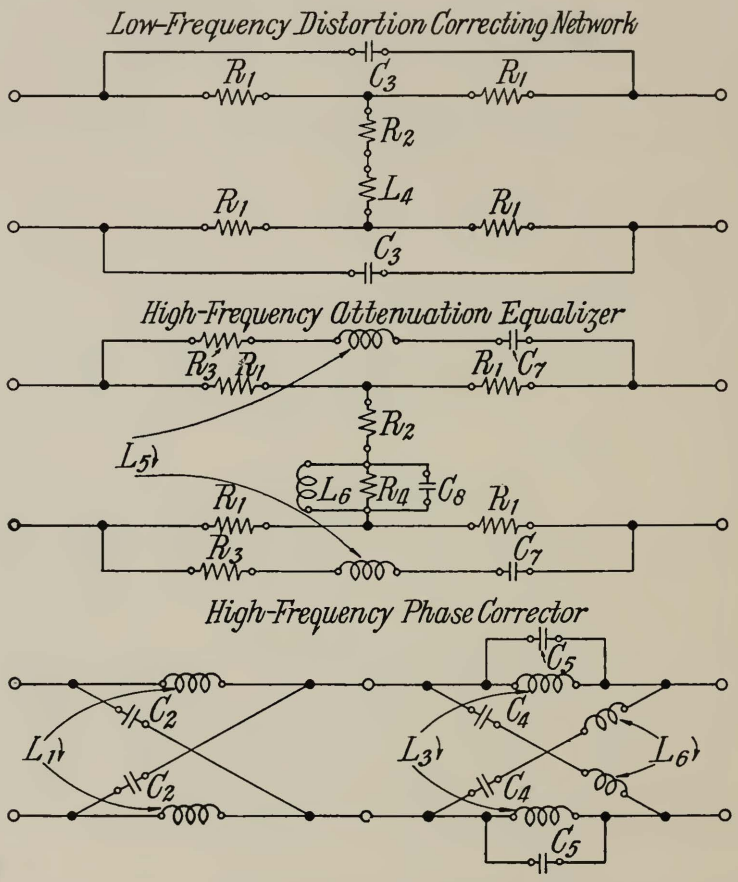


Fig. 16—Distortion correcting networks for television circuit. (Dry weather.)

Since the general circuit arrangement is here the same as in the previous problem, formula (69) is directly applicable. The transfer exponents, $a + ib$, were calculated from it for two weather conditions of the open-wire line, called dry weather and average-wet weather.

[15] "Wire Transmission System for Television," D. K. Gannett and E. I. Green, *B. S. T. J.*, October, 1927, pp. 616–632. See also "The Production and Utilization of Television Signals," Frank Gray, J. W. Horton, and R. C. Mathes, pp. 560–603. Three other papers on Television by H. E. Ives, by H. M. Stoller and E. R. Morton, and by E. L. Nelson are given in the same number of the *Journal*.

From a study of these characteristics and those of certain distortion correcting networks it appeared possible to obtain a base network design for the dry weather condition and supplementary networks for various degrees of wet weather.

The dry weather network consists of three parts in tandem, a low-frequency distortion correcting network, a high-frequency attenuation equalizer, and a high-frequency phase corrector, which were designed in the order given. The final structures are shown in Fig. 16, the first two being put in the form of balanced bridged-T (I*a*) types. The low-frequency distortion correcting section corresponds to Network 1*b*, Appendix IV, and, while designed to approximately equalize the attenuation at low frequencies, it gave at the same time sufficient phase correction in that frequency range. The attenuation data used were

$$f_1 = 50\sim, \qquad A_1 = .409 \text{ napier};$$
$$f_2 = 500\sim, \qquad A_2 = .060 \text{ napier};$$

from which, where $R = 600$ ohms,

$$P_0 = 60{,}307; \qquad Q_0 = 25{,}217;$$
$$a_0 = .2145; \qquad b_1 = 4.945 \cdot 10^{-3};$$
$$R_{11} = 257.4 \text{ ohms}; \qquad C_{12} = 3.056 \text{ mf}.$$

Transformation in the usual manner to the bridged-T (I*a*) type, letting $c = 1/a_0 = 4.662$, gave the balanced structure of Fig. 16 in which

$$R_1 = 64.35 \text{ ohms}; \qquad R_2 = 1334 \text{ ohms};$$
$$C_3 = 6.112 \text{ mf}.; \qquad L_4 = 1.100 \text{ h}.$$

The high-frequency attenuation equalizer was derived from Network 8, Appendix IV, with this data, which followed formula (8) and allowed for the attenuation of the preceding network. The amount of attenuation at the highest frequency was arbitrarily assumed to be .400 napier.

$$f_0 = 0, \qquad A_0 = 2.551 \text{ napiers};$$
$$f_1 = 5{,}000\sim, \qquad A_1 = 2.100 \text{ napiers};$$
$$f_2 = 10{,}000\sim, \qquad A_2 = 1.476 \text{ napiers};$$
$$f_3 = 20{,}000\sim, \qquad A_3 = .400 \text{ napier}.$$

Solution of the linear equations gave

$$P_2 = -53.683 \cdot 10^{-8}; \qquad Q_2 = 5.0669 \cdot 10^{-8}; \qquad Q_4 = 3.7662 \cdot 10^{-18};$$

whence

$$a_0 = .85529; \qquad a_1 = 6.1045 \cdot 10^{-6};$$
$$b_1 = 39.906 \cdot 10^{-6}; \qquad b_2 = 1.9407 \cdot 10^{-9}.$$

Then in the lattice structure

$$R_{11} = 223.8 \text{ ohms}; \qquad L_{12} = 9.668 \text{ mh.};$$
$$C_{13} = .005081 \text{ mf.}; \qquad R_{14} = 1026.3 \text{ ohms}.$$

Transformation to the bridged-T (Ia) type, using as in (70) $c = 1/a_0 = 1.1692$, gave as the elements of the balanced structure of Fig. 16

$$R_1 = 256.6 \text{ ohms}; \qquad R_2 = 94.20 \text{ ohms};$$
$$R_3 = 111.9 \text{ ohms}; \qquad R_4 = 1609 \text{ ohms};$$
$$L_5 = 9.668 \text{ mh.}; \qquad L_6 = 1.829 \text{ mh.};$$
$$C_7 = .010162 \text{ mf.}; \qquad C_8 = .02686 \text{ mf.}$$

Having equalized the dry weather attenuation over the desired frequency range from 10 to 20,000 cycles per second and improved phase conditions at low frequencies, there remained the problem of total phase correction at the higher frequencies. It was found that the high-frequency attenuation equalizer introduced phase distortion at the higher frequencies which was of the same nature but more than twice as great as that due to the original circuit itself. Letting D be the departure from linearity to the value at 20,000 cycles per second of the total phase due to the circuit and the two networks above, the departures at three important frequencies were

$$f_1 = 5{,}000\sim, \qquad D_1 = - .686 \text{ radian};$$
$$f_2 = 10{,}000\sim, \qquad D_2 = - 1.053 \text{ radians};$$
$$f_3 = 20{,}000\sim, \qquad D_3 = 0.$$

A phase characteristic which when combined with these departures can give an approximate linear resultant phase in that frequency range is that of the composite phase Network 16, Appendix IV, containing three parameters. Its phase constant B was therefore taken to satisfy at these three frequencies the relation $B + D = Cf$, or explicitly

$$B = Cf - D. \tag{71}$$

The constant C was arbitrarily chosen so that the network became physical and satisfactory results were given at intermediate frequencies also. After a number of trials the final value taken was $C = .370 \cdot 10^{-3}$.

This then gave

$$M_1 = 2.8015 \cdot 10^{-4}; \qquad M_3 = -1.3929 \cdot 10^{-12}; \qquad N_2 = -2.4671 \cdot 10^{-8}.$$

Whence

$$a_1 = 1.8846 \cdot 10^{-4}; \qquad a_1' = .9169 \cdot 10^{-4}; \qquad b_2' = .7391 \cdot 10^{-8}.$$

These gave as the elements of the high-frequency phase corrector of Fig. 16

$$L_1 = 35.99 \text{ mh.}; \qquad C_2 = .04999 \text{ mf.};$$
$$L_3 = 17.51 \text{ mh.}; \qquad C_4 = .02432 \text{ mf.};$$
$$C_5 = .02138 \text{ mf.}; \qquad L_6 = 15.40 \text{ mh.}$$

The small attenuation effects of coil dissipation were not included in this design and they were later found to be negligible. An equivalent network, namely, Network 15, Appendix IV, might have been used instead of Network 16 for this phase corrector. But since it gives less uniform or practical magnitudes for the inductances and capacities, it would not be the simplest network to construct.

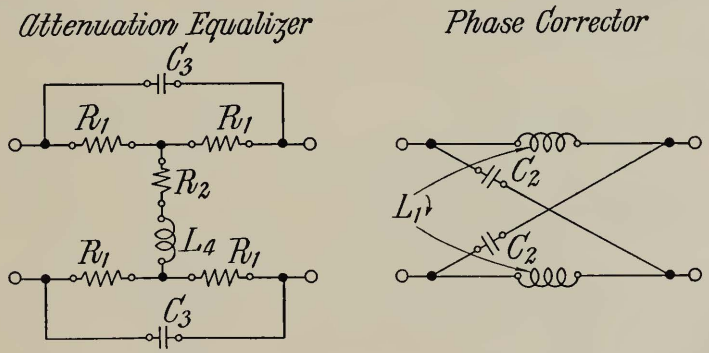


Fig. 17—Weather change distortion correcting networks for television circuit. (One step.)

To provide for wet weather effects on the open-wire part of the circuit, three identical weather change networks were designed each of which was capable of correcting one half the increase in circuit distortion caused by a change from dry weather to average-wet weather. With 0, 1, 2, or 3 of these supplementary networks added in tandem to the dry weather network, provision was thus made for a total of four assumed weather conditions which for convenience I have designated dry, semi-wet, wet, and extra-wet weather. These conditions differed by small equal steps and their number was later found to be

sufficient to cover the weather range ordinarily experienced. The increase, $u + iv$, in the transfer exponent due to a change of one step, such as from dry to semi-wet, was taken as one half the difference of these exponents computed for dry and average-wet weather conditions under which the circuit constants were known. Thus

$$u + iv = \tfrac{1}{2}[(a + ib)_{\text{av.-wet}} - (a + ib)_{\text{dry}}]. \tag{72}$$

The weather change network which corrected this consists of two parts shown in Fig. 17, an attenuation equalizer and a phase corrector, the latter being required primarily because of the phase constant necessarily introduced by the former.

This attenuation equalizer has the same form as the low-frequency network for dry weather and was designed similarly from the data (according to (8))

$$f_1 = 0, \qquad A_1 = .466 \text{ napier};$$
$$f_2 = 20{,}000\sim, \qquad A_2 = .150 \text{ napier}.$$

The assumption for the network of .150 napier at 20,000 cycles per second was found to result in a satisfactory attenuation characteristic over the entire frequency range. Then

$$P_0 = 29{,}872 \cdot 10^4; \qquad Q_0 = 11{,}763 \cdot 10^4;$$
$$a_0 = .22887; \qquad b_1 = .71099 \cdot 10^{-4};$$
$$R_{11} = 274.62 \text{ ohms}; \qquad C_{12} = .04120 \text{ mf}.$$

Transforming to the bridged-T (Ia) structure, $c = 1/a_0 = 4.369$ and the elements of Fig. 17 become

$$R_1 = 68.65 \text{ ohms}; \qquad R_2 = 1242 \text{ ohms};$$
$$C_3 = .08240 \text{ mf}.; \qquad L_4 = 14.83 \text{ mh}.$$

The phase corrector was Network 13, Appendix IV, designed in a manner somewhat different from that usually employed. If D again represents the phase departure of the uncorrected phase from linearity to the value at 20,000 cycles per second, it was found that

$$\text{at } f_1 = 10{,}000\sim, \qquad D_1 = -.111 \text{ radian};$$
$$\text{at } f_2 = 20{,}000\sim, \qquad D_2 = 0.$$

To give a satisfactory resultant phase which is linear through f_1 and f_2 irrespective of its slope, the phase corrector only needed to have a phase constant, B_1 at f_1 and B_2 at f_2, such that

$$D_1 + B_1 = \tfrac{1}{2}B_2, \tag{73}$$

since $f_1 = \frac{1}{2}f_2$. Imposing this condition on the phase relation

$$H = \tan (B/2) = a_1 f, \tag{74}$$

there resulted

$$2 \tan^{-1} (H_2/2) - \tan^{-1} H_2 = - D_1,$$

which can be reduced to

$$\frac{H_2^3}{4 + 3H_2^2} = - \tan D_1,$$

and the cubic in H_2,

$$H_2^3 + 3 \tan D_1 H_2^2 + 4 \tan D_1 = 0. \tag{75}$$

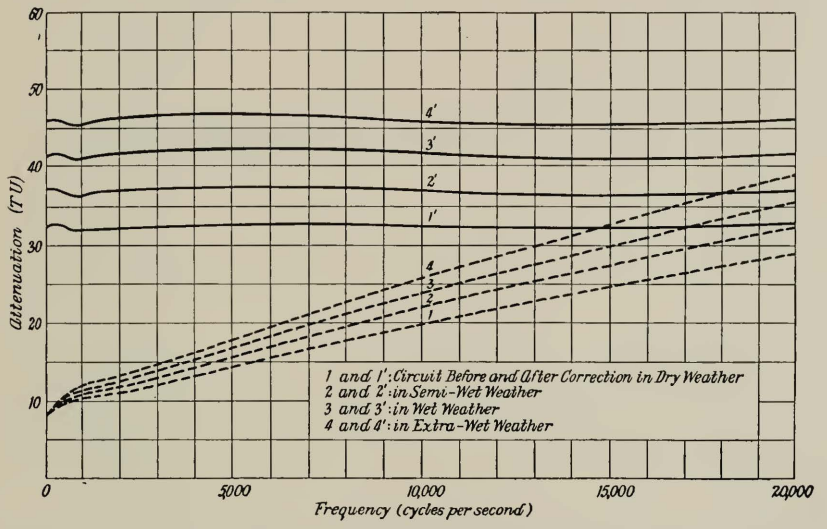


Fig. 18—Attenuation characteristics of television circuit before and after distortion correction.

The solution of the latter with $D_1 = - .111$ radian gave here

$$H_2 = .89363, \qquad \text{whence} \qquad a_1 = H_2/f_2 = .44681 \cdot 10^{-4},$$

and in Fig. 17

$$L_1 = 8.533 \text{ mh.}; \qquad C_2 = .01185 \text{ mf.}$$

For the purpose of showing the amount and precision of distortion correction produced by the addition of these various networks to the open-wire circuit under different weather conditions, attenuation and time-of-phase-transmission characteristics are given in Figs. 18 and 19, respectively. The final results indicate that the design require-

ments were fulfilled. (For measurements and performance of the complete line circuit see footnote 15.)

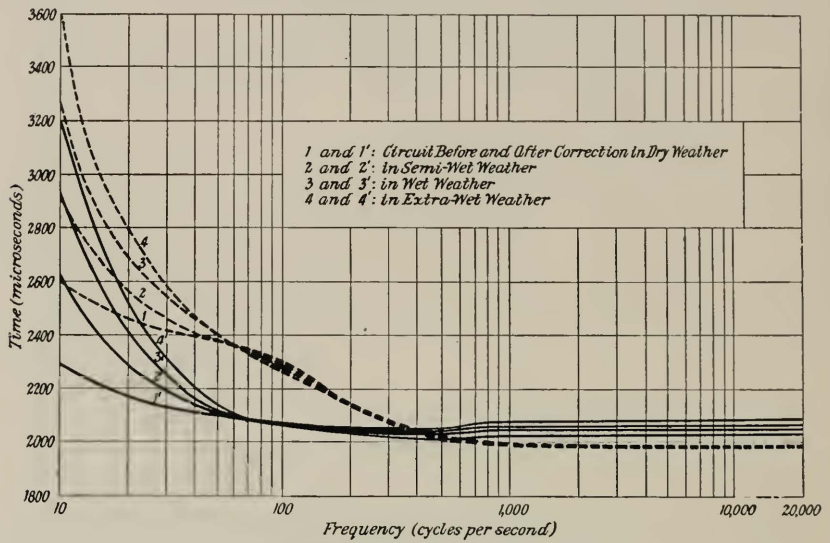


Fig. 19—Time-of-phase-transmission characteristics of television circuit before and after distortion correction.

4.5. *Equalization of Variable Attenuation in Carrier Telephone Circuits*

An open-wire circuit, such as used in a carrier system, is exposed to various weather conditions along the line and consequently experiences considerable changes in its transmission characteristics, primarily its attenuation. For satisfactory operation of carrier circuits the total circuit attenuation must ordinarily be kept reasonably constant.

One practical and advantageous method of maintaining a constant circuit attenuation which takes into account weather changes as well as length differences in the successive repeater sections is the following. Each repeater section is built out and equalized with terminal networks such that at all times the total attenuation has the same uniform value in the desired frequency range. This is done by means of two kinds of networks, *a variable artificial line* and *a base attenuation equalizer*. The variable artificial line builds out any given section to correspond to what is effectively under wet weather conditions the maximum line section used, and the base attenuation equalizer makes this total attenuation of the section uniform in the frequency range under consideration. Then the total attenuation of any line section, artificial line, and attenuation equalizer has the same constant value over the frequency range and will thus be in proper adjustment with a repeater having a fixed gain.

Such an artificial line is made up of a number of different sections whose various tandem combinations can build up by small steps a considerable length of repeater section. A mechanism for switching the various sections of artificial line in and out of a repeater section might be operated by means of regulating apparatus which is automatically controlled from circuit conditions existing on a single-frequency pilot channel or channels. In Fig. 20 is shown a type of network suitable for a section of such artificial line. It is equivalent to Network 3*a*, Appendix IV, from which it can be transformed. The

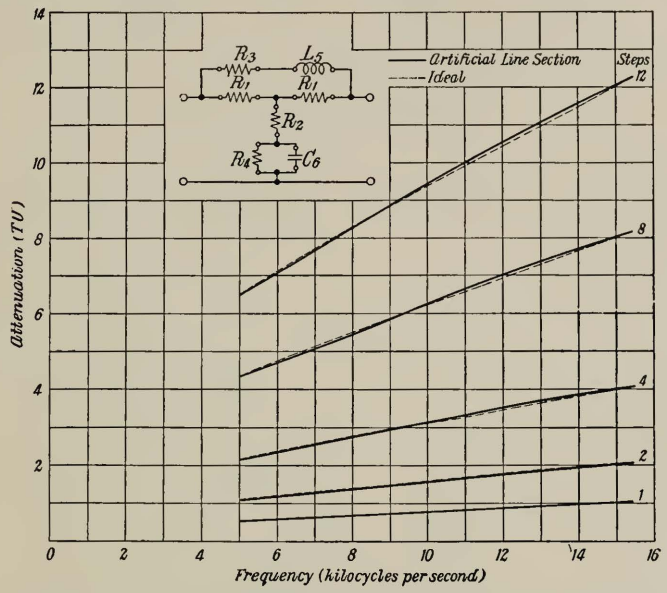


Fig. 20—Sections of variable artificial line and their attenuation characteristics for carrier telephone circuits.

following table gives the network elements for a group of such sections. I need not discuss any of the design details here but shall merely state that these sections were designed according to formulæ in Appendix IV from attenuation data which represent average requirements on the open-wire pairs used for carrier systems. The frequency range for these networks, 5.0–15.4 kilocycles per second, includes a lower group of adjacent carrier channels each having a band width of about 2500 cycles per second.

The attenuation characteristics of these individual sections are also given in Fig. 20. By properly combining them the desired maximum amount of artificial line can be obtained in equal steps, each step corresponding to approximately 1 T.U. at the highest frequency of the range.

TABLE III

ARTIFICIAL LINE CONSTANTS (Fig. 20)

(5.0–15.4 kilocycles per second)

	Steps				
	1	2	4	8	12
R_1 (ohms).....	54.2	108.2	212.9	401.4	605.0
R_2.....	3348.	1637.	753.2	255.3	0
R_3.....	40.2	80.5	160.9	318.4	445.3
R_4.....	9108.	4544.	2275.	1150.	822.1
L_5 (mh.).....	1.62	3.24	6.57	13.83	23.13
C_6 (mf.).....	.004426	.008863	.01794	.03779	.06318

Iterative Impedance R = 605 ohms.

A structure suitable for a base attenuation equalizer is that of Fig. 21, transformed from Network 11, Appendix IV. In designing it to simulate the required attenuation characteristic shown, the procedure

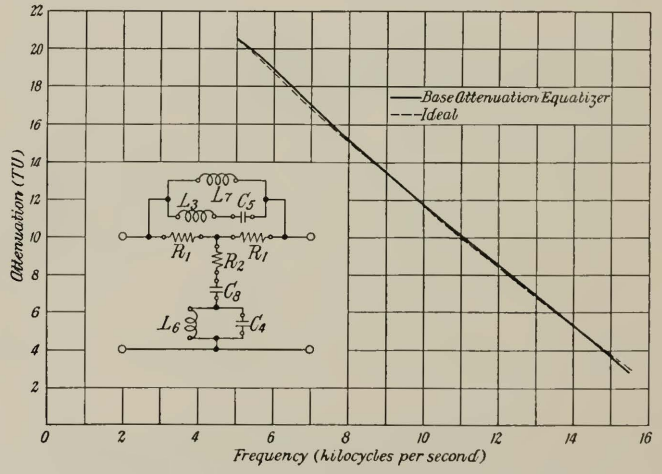


Fig. 21—Base attenuation equalizer and its attenuation characteristic for carrier telephone circuits.

was first to choose arbitrarily a plausible maximum attenuation for the network and then to use in the attenuation linear equations the three desired attenuation values at the mean and the extreme frequencies of the frequency range. At the highest frequency the attenuation was lowered slightly to allow for coil dissipation. Several such computations were made with different values of this maximum until a network was derived which gave a satisfactory result at all frequencies

within the range. The magnitudes of the elements corresponding to the partial attenuation characteristic shown in Fig. 21, where $R = 605$ ohms, are

$R_1 = 508.4$ ohms; $R_2 = 105.8$ ohms;
$L_3 = 12.69$ mh.; $C_4 = .03469$ mf.;
$C_5 = .005852$ mf.; $L_6 = 2.14$ mh.;
$L_7 = 238.8$ mh.; $C_8 = .6525$ mf.

The departures of the attenuation from the desired values are less than .2 T.U. At the highest frequencies small coil dissipation tends to improve this result.

4.6. *Phase Correction in Transatlantic Telephone System*

At the receiving stations of the transatlantic telephone system it is necessary to use phase correctors in connection with the antenna arrays. These networks serve in two capacities, either (*a*) as artificial lines or delay networks which build out the phase characteristics of short transmission lines until they are equivalent to certain longer lines used elsewhere in the system, or (*b*) as phase correctors which secure adjustable and arbitrary phase characteristics when combining the outputs of the antennæ which form the array. For satisfactory operation the phase correctors had to meet these design requirements.

1. A constant iterative impedance of $R = 600$ ohms.
2. A continuously variable phase change which is proportional to frequency over the frequency range from 50 to 65 kilocycles per second, the total phase change being from 0 to 250 degrees at 50 kilocycles per second.
3. Over any frequency band of 5 kilocycles per second in the range the variations should be less than .100 degree for the phase and less than .025 T.U. for the attenuation.
4. A balanced structure.

In making the design it was found that the continuously variable phase change to the desired maximum could be provided by means of one variable section having a small phase constant and five fixed sections of Networks 13, 14, and 16, Appendix IV. Designated in terms of their phase constants at 50 kilocycles per second as in Fig. 22, the variable section has a range of 0–15 degrees, while the fixed sections have phase constants of 10, 20, 40, 80, and 160 degrees, respectively. The variable section is normally required to give a maximum of only 10 degrees but an extension of its range to 15 degrees is provided so as to ensure phase overlapping at any transition point where a

section is put in or taken out of the circuit. By properly combining these sections it is seen that a continuous range from 0 to 325 degrees is obtainable.

The sections were designed from the formulæ of Appendix IV so as to give the desired individual linear phase characteristics shown in Fig. 22. It need only be stated that the data taken from the phase

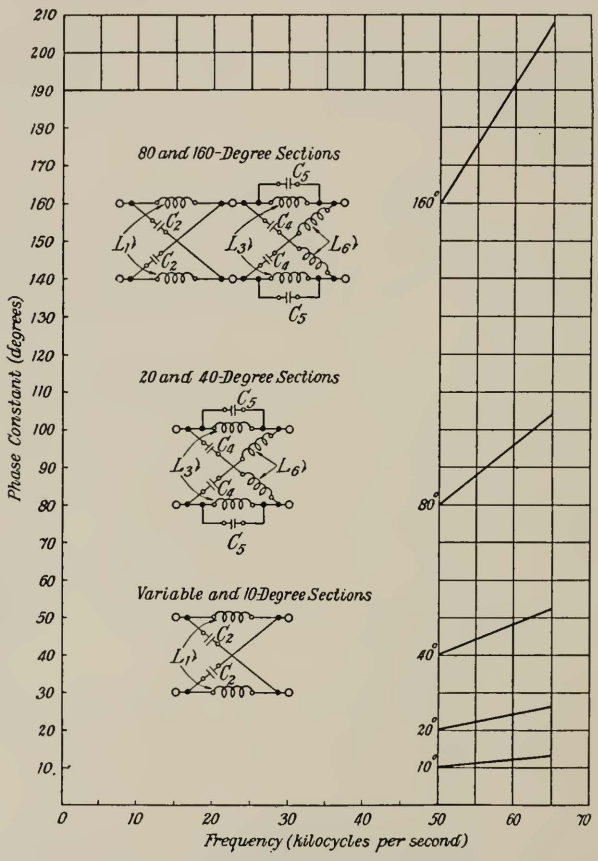


Fig. 22—Sections of variable phase corrector and their phase characteristics used in the transatlantic telephone system.

characteristics in the one-parameter sections were those at 50, in the two-parameter sections those at 50 and 65, and in the three-parameter composite sections those at 50, 57.5, and 65 kilocycles per second. The elements for the *variable section* in Fig. 22 are continuously variable and have their magnitudes given in terms of the variable phase constant B at 50 kilocycles per second as

$$L_1 = \frac{\tan (B/2)R}{50\pi} \text{ mh.}; \qquad C_2 = \frac{10 \tan (B/2)}{\pi R} \text{ mf.}$$

The results for the fixed sections follow in Table IV.

TABLE IV

PHASE CORRECTOR CONSTANTS

(Fig. 22)

	Fixed Sections (Degrees at 50 kilocycles per second)				
	10	20	40	80	160
L_1 (mh.)	.334			1.211	2.941
C_2 (10^{-3} mf.)	.464			1.682	4.084
L_3 (mh.)		.667	1.333	1.456	2.466
C_4 (10^{-3} mf.)		.926	1.851	2.023	3.425
C_5 (10^{-3} mf.)		.309	.631	1.051	2.408
L_6 (mh.)		.223	.454	.756	1.734

In any one of these sections the computed departures of the phase constant from ideal proportionality to frequency in the frequency range 50 to 65 kilocycles per second was usually much less than .02 degree. The practical construction of the networks gave similar high precision, and by using coils of small dissipation constant, d = (resistance/reactance), the attenuation requirements were likewise satisfied. The frequency band now in use is from 58.5 to 61.5 kilocycles per second.

It may be added that these designs can readily be altered so as to apply to other frequency ranges. In order to translate the phase constants from the 50-kilocycle designation to any other frequency range with a minimum frequency, f_0, designation, multiply all inductances and capacities by the translation factor $(50{,}000/f_0)$.[16]

4.7. *Simulation of Smooth Line*

This application is based upon and illustrates the general results of Part 3 which discusses recurrent networks having arbitrary iterative impedances. A network design will be given which has the following characteristics.

1. A propagation constant which simulates a moderate propagation length of any smooth line, or its equivalent.

[16] For a discussion of other applications of constant resistance networks see footnote 10; also "Transmission Circuits for Telephonic Communication," K. S. Johnson.

2. An iterative impedance which equals that of the smooth line at all frequencies.

Such a network could have a number of uses. For example, it could serve as a substitute for a small length of smooth line where approximately exact simulation is required as in certain laboratory tests, or as part of an artificial line in a balancing network. Leakance changes can be provided for by means of particular adjustable resistances. The design can represent the special case of a distortionless line at the lower frequencies and, if non-dissipative, give a phase network having a constant time-of-phase-transmission in this frequency range.

The method of solution differs considerably from those previously used for the other networks and so will be given here. To begin with let

z_a = series, z_b = shunt } impedance of any section of smooth line, or its equivalent, of propagation constant γ, iterative impedance k, and length l.

Also let

X = open-circuit, Y = short-circuit } impedance of the smooth line section.

Then

$$\gamma l = \sqrt{z_a/z_b} = \tanh^{-1} \sqrt{Y/X},$$

and

$$k = \sqrt{z_a z_b} = \sqrt{XY}. \tag{76}$$

(*B. S. T. J.*, October, 1924, p. 617.)

From these

$$z_a = k\gamma l = \sqrt{XY} \tanh^{-1} \sqrt{Y/X},$$

and

$$z_b = k/\gamma l = \sqrt{XY}/\tanh^{-1} \sqrt{Y/X}; \tag{77}$$

thus z_a and z_b are inverse networks of impedance product k^2. In a physical smooth line z_a is simulated by series resistance and inductance and z_b by parallel resistance and capacity (assuming the line constants to be independent of frequency), both represented by simple physical networks. In other cases they may be realized in desired frequency ranges, more or less approximately, by physical networks. It will be assumed in what follows that z_a and z_b are given by the above formulæ.

The structure which is to simulate the smooth line is shown in its general form as Network 18, Appendix IV, wherein z_a and z_b are considered as two types of physical elements whose combinations in

different proportions make up the network. It consists of a composite lattice network of two sections having four real, positive parameters, m_1, m_2, m_1', and m_2', two in each section.

In the first section put for the series impedance

$$z_{11} = \frac{1}{\frac{1}{2m_1 z_a} + \frac{1}{2z_b/m_2}}. \tag{78}$$

To satisfy the condition for the desired iterative impedance at all frequencies,

$$K = \sqrt{z_{11}z_{21}} = k = \sqrt{z_a z_b}, \tag{79}$$

it follows that the lattice impedance must be

$$z_{21} = \frac{m_2 z_a}{2} + \frac{z_b}{2m_1}. \tag{80}$$

That is, z_{11} and z_{21} are also inverse networks of impedance product k^2. The propagation constant, by generalized (13) (that is, R replaced by K), is

$$e^{\Gamma} = \frac{1 + m_1 y + m_1 m_2 y^2}{1 - m_1 y + m_1 m_2 y^2}, \tag{81}$$

where for convenience $y = \sqrt{z_a/z_b} = \gamma l$ = propagation length.

In the second section, similarly,

$$z_{11}' = \frac{1}{\frac{1}{2m_1' z_a} + \frac{1}{2z_b/m_2'}},$$

$$z_{21}' = \frac{m_2' z_a}{2} + \frac{z_b}{2m_1'}, \tag{82}$$

and

$$e^{\Gamma'} = \frac{1 + m_1' y + m_1' m_2' y^2}{1 - m_1' y + m_1' m_2' y^2}.$$

For the composite structure made up of these two sections in tandem, the iterative impedance condition is already fulfilled independently of the values of the coefficients, since (79) holds for each section. Its propagation constant is given from (81) and (82) by

$$e^{\Gamma_e} = e^{\Gamma+\Gamma'}$$

$$= \frac{\begin{array}{l}1 + (m_1 + m_1')y + (m_1 m_2 + m_1 m_1' + m_1' m_2')y^2 \\ \quad + (m_1 m_2 m_1' + m_1 m_1' m_2')y^3 + m_1 m_2 m_1' m_2' y^4\end{array}}{\begin{array}{l}1 - (m_1 + m_1')y + (m_1 m_2 + m_1 m_1' + m_1' m_2')y^2 \\ \quad - (m_1 m_2 m_1' + m_1 m_1' m_2')y^3 + m_1 m_2 m_1' m_2' y^4\end{array}}. \tag{83}$$

It remains to choose the coefficients m_1, m_2, m_1', and m_2' so that for moderate propagation lengths, $y = \gamma l$, the composite network will give

$$\Gamma_c \text{ approximately } = y = \gamma l. \tag{84}$$

At this point let us introduce an important simplification by writing the function

$$e^y = \frac{e^{y/2}}{e^{-y/2}} = \frac{1 + \frac{1}{2}y + \frac{1}{8}y^2 + \frac{1}{48}y^3 + \frac{1}{384}y^4 + \frac{1}{3840}y^5 + \cdots}{1 - \frac{1}{2}y + \frac{1}{8}y^2 - \frac{1}{48}y^3 + \frac{1}{384}y^4 - \frac{1}{3840}y^5 + \cdots}. \tag{85}$$

Then upon comparing (83) and (85) we see that fortunately for small values of y we can satisfy (84) providing we identify the coefficients of powers of y in (83) as

$$\begin{aligned} m_1 + m_1' &= \frac{1}{2}, \\ m_1 m_2 + m_1 m_1' + m_1' m_2' &= \frac{1}{8}, \\ m_1 m_2 m_1' + m_1 m_1' m_2' &= \frac{1}{48}, \end{aligned} \tag{86}$$

and

$$m_1 m_2 m_1' m_2' = \frac{1}{384}.$$

The solution of the equations gives a sixth degree equation for m_1, namely,

$$m_1^6 - \frac{3}{2}m_1^5 + m_1^4 - \frac{3}{8}m_1^3 + \frac{5}{64}m_1^2 - \frac{1}{128}m_1 + \frac{1}{4608} = 0;$$

and for the others

$$m_2 = \frac{6m_1 - 48m_1^2 m_1' - 1}{48m_1(m_1 - m_1')},$$

$$m_1' = .5 - m_1,$$

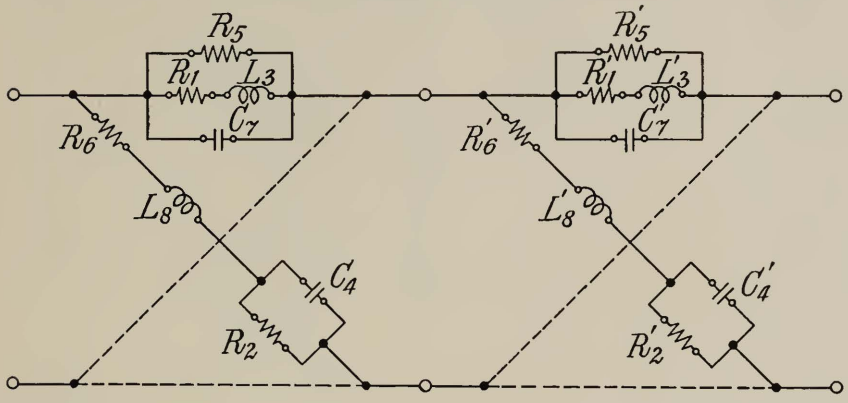
and

$$m_2' = \frac{1 + 48m_1(m_1')^2 - 6m_1'}{48m_1'(m_1 - m_1')}.$$

From these we get this set of real positive coefficients, determined once for all, namely,

$$\begin{aligned} m_1 &= .45737; \qquad m_2 = .14456; \\ m_1' &= .04263; \qquad m_2' = .92403. \end{aligned} \tag{87}$$

With the above fixed values of the coefficients and formulæ (77), (78), (80), and (82), the network can be constructed which is to simulate any smooth line having physically realizable z_a and z_b. This simula-



(Broken lines indicate the other series and lattice branches, respectively identical.)

$R_1 = m_1R'l,$	$R_2 = 1/m_1G'l,$	$R_1' = m_1'R'l,$	$R_2' = 1/m_1'G'l,$
$L_3 = m_1L'l,$	$C_4 = m_1C'l,$	$L_3' = m_1'L'l,$	$C_4' = m_1'C'l,$
$R_5 = 1/m_2G'l,$	$R_6 = m_2R'l$	$R_5' = 1/m_2'G'l,$	$R_6' = m_2'R'l,$
$C_7 = m_2C'l,$	$L_8 = m_2L'l,$	$C_7' = m_2'C'l,$	$L_8' = m_2'L'l,$
$m_1 = .45737,$	$m_2 = .14456,$	$m_1' = .04263,$	$m_2' = .92403.$

Fig. 23—Artificial smooth line which simulates a moderate length, l, of line having the primary constants R', L', G', and C' per unit length. (If $R' = G' = 0$, it becomes a non-dissipative phase network whose time-of-phase-transmission at the lower frequencies has the constant value, $\tau_p = \sqrt{L'C'}l$.)

tion is very accurate for small values of y. As y increases, the departure of the network propagation characteristic from the smooth line values also increases, but it amounts to less than 1.4 per cent even at $|y| = 3.0$, as may be derived from a comparison of (83) and (85).

As an illustration of this type of design, these results were analytically applied to the case of a 104-mil open-wire smooth line having the constants per loop mile (for wet weather, and assumed independent of frequency),

$R' = 10.12$ ohms; $\quad L' = 3.66$ mh.;
$G' = 3.20$ micromhos; $\quad C' = .00837$ mf.

The corresponding simulating network for a length l is shown structurally in Fig. 23, where

and

$$z_a = (R' + iL'\omega)l,$$
$$z_b = 1/(G' + iC'\omega)l. \tag{88}$$

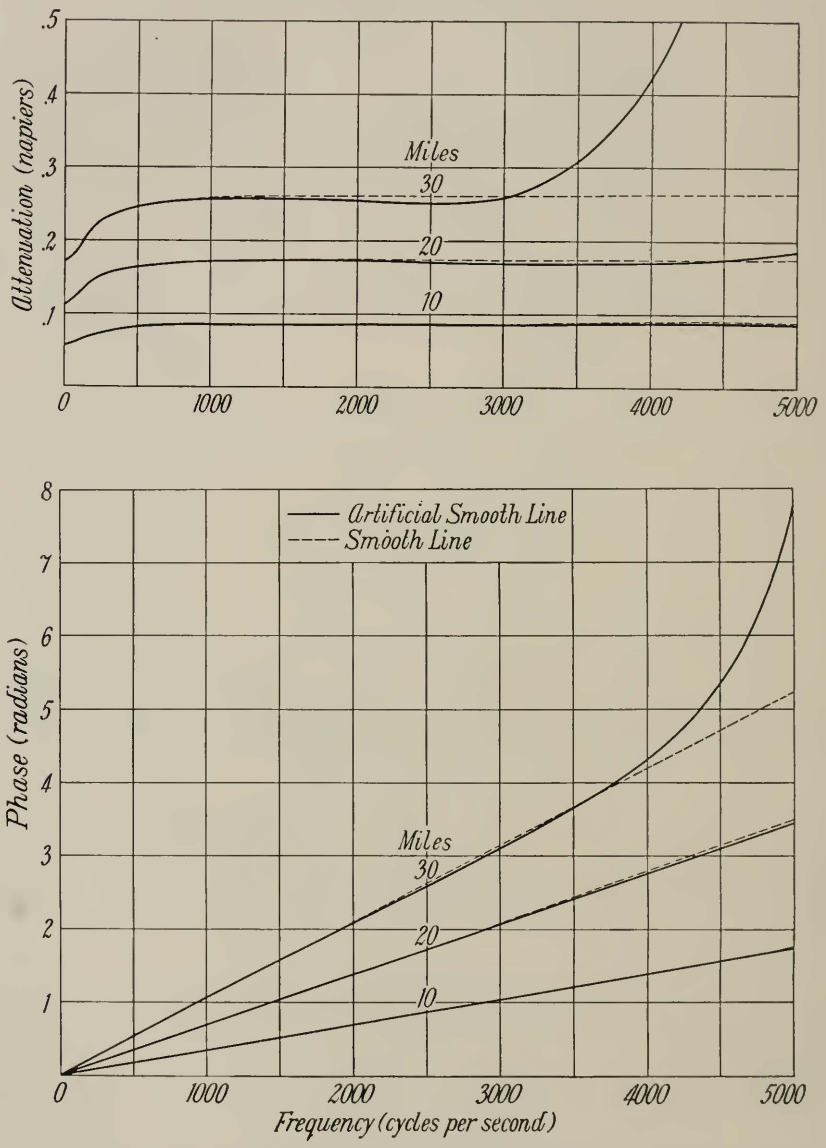


Fig. 24—Propagation characteristics of 10-, 20- and 30-mile lengths of 104-mil open-wire smooth line and of the simulating artificial smooth lines. (R' = 10.12 ohms, L' = 3.66 mh., G' = 3.20 micromhos (wet weather), and C' = .00837 mf. per loop mile.)

A comparison of the propagation characteristic of a line section and that of its simulating network is shown in Fig. 24 for three different line lengths, $l = 10$, 20, and 30 miles. Even in the longest section the simulation is good up to 3000 cycles per second. The iterative impedances are, of course, identical as in Fig. 25.

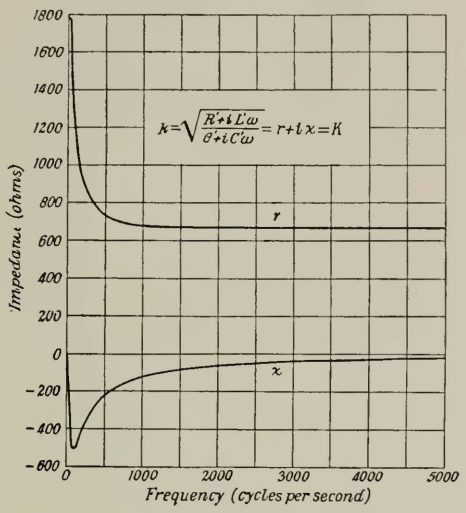


Fig. 25—Iterative impedance of 104-mil open-wire smooth line and of the simulating artificial smooth lines.

While the above general design considered four parameters, a similar procedure can be followed with other networks having a smaller or greater number of parameters. The structure can be obtained by building the series impedance of any section out of various combinations of the impedance elements z_a and z_b. However, several of the above four-parameter composite sections can perhaps meet most design requirements.

Appendix I

Discussion of Linear Phase Intercept

Let us first consider steady-state transmission over a circuit where the impressed e.m.f., consisting of simple sinusoids of any two angular frequencies ω_1 and ω_2, is given by

$$\begin{aligned} E(t) &= \sin \omega_1 t + \sin \omega_2 t, \\ &= 2 \cos \tfrac{1}{2}(\omega_1 - \omega_2)t \sin \tfrac{1}{2}(\omega_1 + \omega_2)t. \end{aligned} \tag{89}$$

Assume that the circuit has at these frequencies the transfer exponents $a_1 + ib_1$ and $a_2 + ib_2$ such that $a_1 = a_2 = a'$. A straight line drawn

through b_1 and b_2 in the ω, b plane will have a slope τ, say, and at $\omega = 0$ a linear phase intercept b_0 which may have any value. Hence, the transfer exponent may be expressed as a function of frequency at these two frequencies by the relations

$$\begin{aligned} a &= a' = \text{constant}, \\ \text{and} \qquad & \\ b &= \tau\omega + b_0. \end{aligned} \tag{90}$$

The received voltage across R will then be a periodic function which is attenuated by an amount a' napiers and is

$$\begin{aligned} v(t) &= e^{-a'}\left[\sin\left(\omega_1(t-\tau) - b_0\right) + \sin\left(\omega_2(t-\tau) - b_0\right)\right], \\ &= 2e^{-a'}\cos\tfrac{1}{2}(\omega_2 - \omega_1)(t-\tau)\sin\left(\tfrac{1}{2}(\omega_1 + \omega_2)(t-\tau) - b_0\right). \end{aligned} \tag{91}$$

How the transmitting property of this circuit for the two frequencies depends upon the phase intercept can be seen from a comparison of (91) with (89). In order that the received voltage may be a time-function of identically the same shape as the impressed voltage, but with a time-of-transmission over the circuit of τ seconds, it is necessary that $b_0 = 2n\pi$ radians, where n is any positive or negative integer. This would mean no distortion of the impressed steady-state signal made up of the two frequency components. If $b_0 = (2n \pm 1)\pi$, there would be an apparent distortion only of a reversal in sign. However, if $b_0 = (2n \pm \tfrac{1}{2})\pi$, there would be maximum distortion in the transmitted voltage. These conclusions may be tabulated briefly as follows:

If $b_0 = 2n\pi$, no distortion;

If $b_0 = (2n \pm 1)\pi$, apparent distortion of sign reversal;

If $b_0 = (2n \pm \tfrac{1}{2})\pi$, maximum distortion.

The above discussion considered the case of any two frequencies. If now we assume that the circuit has the characteristics (90) for several or a range of frequencies, then the conclusions above obviously apply as well to the steady-state transmission of an impressed e.m.f. which is made up of any of those frequencies. Thus, *for distortionless steady-state transmission (without change of signal shape), the transfer exponent must have for the frequency components impressed not only a uniform attenuation and a linear phase relation, but also a proper linear phase intercept* $b_0 = 2n\pi$. If, in a physical system, (90) is satisfied over a frequency range which includes zero frequency, then τ would necessarily be positive and $b_0 = 0$ or a multiple of 2π.

Proceeding next to the transmission of an e.m.f. impressed suddenly

at time $t = 0$, we note that since the e.m.f. can be expressed in terms of a Fourier integral representation from $t = -\infty$ to $t = +\infty$ we may regard it as made up of a distribution of steady-state components. For example, let the e.m.f. of (89) be impressed on a circuit at time $t = 0$. Then

$$\begin{aligned} E(t) &= \left(\frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\sin ty}{y}\,dy\right)(\sin \omega_1 t + \sin \omega_2 t), \\ &= \tfrac{1}{2}(\sin \omega_1 t + \sin \omega_2 t) + \frac{1}{\pi}\int_0^\infty \left(\frac{\omega_1}{\omega_1^2 - \omega^2} + \frac{\omega_2}{\omega_2^2 - \omega^2}\right)\cos t\omega\, d\omega. \end{aligned} \tag{92}$$

This represents the impressed voltage for negative as well as positive values of t since in the first equation the factor of the sinusoids represents a function which is zero for all negative and unity for all positive values of t. We may then interpret the last equation as giving for all values of time the frequency distribution of steady-state components of all frequencies which give the same result as the sinusoids of (89) impressed suddenly at $t = 0$. This distribution extends over the entire frequency range and has the largest amplitudes about $\omega = \omega_1$ and $\omega = \omega_2$.

Hence, if the initial part of the impressed e.m.f., as well as the final steady state, is to be transmitted without distortion, the circuit transfer voltage must have a characteristic which is distortionless not only with respect to ω_1 and ω_2 but also to all angular frequencies about them as obtained from the analysis. That is, since the steady state is only the limiting case of the transient state, an ideal circuit characteristic for its distortionless transmission is only a part of and is included in that for the transient state. Or, vice versa, ideal circuit characteristics for the steady state are at least the same as for the transient state.

These results are useful in studying a circuit whose attenuation is constant and whose phase characteristic is approximately linear over an internal frequency band. An extrapolation of this linear phase characteristic to zero frequency may give a phase intercept which is not ideal for preservation of wave-shape even in the steady state of frequencies within the band, as we have seen. Increasing the frequency range over which an ideal phase relation holds obviously improves the transmission of transient voltages. Practically, good results are obtained in a circuit wherein the attenuation is approximately constant and the phase is approximately proportional to frequency over the required internal band of frequencies; then the phase intercept, b_0, is zero and the time-of-phase-transmission, $\tau_p = b/\omega$, is approximately constant and represents the time-of-transmission of the circuit for those frequencies.

Appendix II

Proofs of Linear Transducer Theorems

Theorem I: Any passive network whose attenuation constant is zero at all frequencies is a limiting case of a physical wave-filter wherein the transmitting band extends over the entire frequency range. The proof that the phase constant increases with frequency in the transmitting band of any wave-filter has already been given by the writer in the paper, "Theory and Design of Uniform and Composite Electric Wave-Filters," *B. S. T. J.*, January, 1923, pages 37–38. In the present case, therefore, the phase constant increases throughout the frequency range.

The proof relating to the iterative impedance will be given in two steps which comprise essentially the proofs of two impedance theorems. From the first of these it will follow immediately that the transducer under consideration has everywhere a real iterative impedance because of symmetry and a transmitting band extending over the entire frequency range; from the second, this real iterative impedance is a constant resistance throughout the frequency range.

Wave-Filter Impedance Theorem: In all transmitting bands the iterative impedances of a recurrent section of any electric wave-filter are conjugate impedances. If the section is symmetrical, they are equal and real without a reactance component.

From the general formulæ on page 617 of *B. S. T. J.*, October, 1924, we may write the iterative impedances as:

$$\left.\begin{matrix} K_a \\ K_b \end{matrix}\right\} = \tfrac{1}{2}((X_a + X_b) \tanh \Gamma \pm (X_a - X_b)), \qquad (93)$$

where X_a and X_b are the open-circuit driving-point impedances at the ends a and b of the transducer. In a wave-filter recurrent section which is made up of non-dissipative reactance elements the impedances X_a and X_b have only reactance components. Also, in a transmitting band the attenuation constant is zero, so that here $\Gamma = iB$ and $\tanh \Gamma = i \tan B$. From this, it follows readily that in any transmitting band the first term of the right-hand member of (93) represents a positive resistance component and the second term a reactance component. Hence, the resistance components of K_a and K_b are identical while their reactance components differ only in sign; that is, K_a and K_b are conjugate impedances in all transmitting bands.

As results of the above we may state parenthetically:

Corollary I: The absolute values of the iterative impedances of a wave-filter recurrent section are equal at any frequency in all transmitting bands; and

Corollary II: The iterative impedances of a wave-filter recurrent section are such as to give maximum energy transfer from section to section in all transmitting bands.

When the section is symmetrical, $X_a = X_b$, and therefore $K_a = K_b = r$, a resistance in those frequency ranges.

Non-Reactive Impedance Theorem: The impedance of any two-terminal network whose reactance component is zero at all frequencies must have a resistance component which is constant, independent of frequency. To prove the theorem, let the impedance of any two-terminal network whose reactance component is zero at all frequencies be represented as:

$$Z = r, \tag{94}$$

where r is a real function of frequency.

The general relations between the components of the steady-state admittance, $\alpha(\omega) + i\beta(\omega)$, of a network and the corresponding indicial admittance, $h(t)$, are known from electric circuit theory to be:

$$\alpha(\omega) = h(o) + \int_0^\infty \cos \omega y h'(y) dy$$

and

$$\beta(\omega) = -\int_0^\infty \sin \omega y h'(y) dy; \tag{95}$$

also

$$h(t) = \alpha(o) + \frac{2}{\pi}\int_0^\infty \frac{\beta(\omega)}{\omega} \cos t\omega d\omega, \quad t > 0.$$

(See pages 18 and 180 of the reference in footnote 5.)

In the passive network under discussion here, the admittance components at all frequencies from (94) are

$$\alpha(\omega) = 1/r, \tag{96}$$

and

$$\beta(\omega) = 0.$$

Upon substituting them in (95) it is found that

$$h(t) = \alpha(o) = \text{a constant}, \quad t > 0, \tag{97}$$

$$h'(t) = 0$$

and

$$\alpha(\omega) = 1/r = h(o) = \text{a constant}.$$

This relation demands that the resistance component r be constant, independent of frequency, as stated in the theorem.

The converse of the above theorem does not follow, that is, if the resist-

ance component of a two-terminal network impedance is constant, independent of frequency, it is not necessary that the reactance component be zero throughout the frequency range. This may be seen from the relations above. A simple example is series resistance and inductance.

Theorem II: If the iterative impedance of a network is real at all frequencies, it must be constant according to the latter impedance theorem above.

For the second part of Theorem II we have as assumptions regarding the propagation constant, $\Gamma = A + iB$, and iterative impedance, K, effectively

$$B = \tau\omega$$

and (98)

$$K = \text{a constant} = R,$$

where τ is some positive constant. The transfer admittance components with respect to a resistance R which terminates the transducer are then

$$\alpha(\omega) = \frac{e^{-A}}{R}\cos \tau\omega$$

and (99)

$$\beta(\omega) = -\frac{e^{-A}}{R}\sin \tau\omega.$$

By means of these and (95) we shall prove that A is uniform at all frequencies.

To satisfy (95) with (99) at all frequencies the transducer must be such as to give the relations

$$h(o) = o,$$

$$h'(t) = o, \quad t \neq \tau,$$

and (100)

$$\int_{\tau-}^{\tau+} h'(y)dy = \frac{e^{-A}}{R}.$$

Since the left-hand member of the last relation is independent of frequency, it follows necessarily that the attenuation constant, A, must be uniform. That uniform attenuation together with (98) is also sufficient to satisfy the other relations of (100) can be seen if the parameter characteristics at all frequencies are

$$A = \text{a constant},$$

$$B = \tau\omega \tag{101}$$

and

$$K = \text{a constant} = R.$$

From electric circuit theory, the fundamental integral equation for the indicial admittance $h(t)$ becomes

$$\frac{1}{pZ(p)} = \frac{e^{-A-\tau p}}{pR} = \int_0^\infty e^{-py} h(y)dy, \tag{102}$$

where p replaces $i\omega$. Its solution is

$$\begin{aligned} &h(t) = o, \quad t < \tau \\ \text{and} \\ &h(t) = \frac{e^{-A}}{R} = \text{a constant}, \quad t > \tau; \end{aligned} \tag{103}$$

whence also $h'(t) = o$ for $t \neq \tau$, thus satisfying (100). These results hold as well for the limiting case of $B = o$, meaning $\tau = o$.

It may be pointed out here also that *the converse of the latter theorem does not follow.* That is, if the transducer has a uniform attenuation constant and a constant resistance iterative impedance, it is not necessary that the phase constant be proportional to frequency throughout the range. This is seen from the general equations or from the fact that we can alter the phase characteristic non-linearly by means of phase networks having zero attenuation and a constant resistance iterative impedance.

Theorem III: A symmetrical transducer made up entirely of resistances would have the characteristics

$$\begin{aligned} &A = \text{a constant}, \\ &B = o \\ \text{and} \\ &K = \text{a constant} = R. \end{aligned} \tag{104}$$

Many other more complicated networks satisfying (104) are known to exist, as in Section 4.1. We need not, therefore, seek further to prove the possible existence of such a combination of parameters.

For networks in which B is not zero, but

$$\begin{aligned} &A = \text{a constant} \\ \text{and} \\ &K = \text{a constant} = R, \end{aligned} \tag{105}$$

the transfer admittance components with respect to a terminating resistance R are given as

$$\begin{aligned} &\alpha(\omega) = \frac{e^{-A}}{R} \cos B \\ \text{and} \\ &\beta(\omega) = -\frac{e^{-A}}{R} \sin B. \end{aligned} \tag{106}$$

Using these and the general relations (95), we can obtain

$$\frac{dB}{d\omega} = \frac{\int_0^\infty y \sin \omega y h'(y) dy}{\int_0^\infty \sin \omega y h'(y) dy}, \tag{107}$$

which is independent of A.

Since (if B is not everywhere zero) $dB/d\omega$ is positive when $A = o$ according to Theorem I, and since by (107) it is independent of A (a constant), it will be positive whatever the value of A. Hence, B increases with frequency in such transducers.

Appendix III

Propagation Constant and Iterative Impedance Formulæ for General Ladder, Lattice and Bridged-T Types

These formulæ apply to the general types of structures shown in Fig. 2 and should be used whenever it is desired to take into account accurately the actual physical impedances. Network designs which follow the methods given in this paper are made under the assumption of invariable lumped elements. In constructing physical networks according to such designs, however, certain departures from this assumption unavoidably make their appearance and must be taken into consideration whenever extreme accuracy is required. The departures include dissipation in coils and condensers, distributed capacity in coils, as well as inaccuracies due to manufacture.

Some of these formulæ have been given in previous papers but all can be derived readily either by the method given in *B. S. T. J.*, January, 1923, p. 34, or by that in *B. S. T. J.*, October, 1924, p. 617.

Ladder Type:

$$\cosh \Gamma = 1 + \tfrac{1}{2}\frac{z_1}{z_2}. \tag{108}$$

The iterative impedances at different terminations are:

$$\begin{aligned} &\text{At full-series} = K_1 + \tfrac{1}{2}z_1, \\ &\text{At full-shunt} = K_1 - \tfrac{1}{2}z_1, \\ &\text{At mid-series} = K_1 = \sqrt{z_1 z_2 + \tfrac{1}{4}z_1^2}, \\ &\text{At mid-shunt} = K_2 = z_1 z_2 / K_1. \end{aligned} \tag{109}$$

Lattice Type:

$$\cosh \Gamma = 1 + \frac{2z_1}{4z_2 - z_1} \tag{110}$$

and

$$K = \sqrt{z_1 z_2}. \tag{111}$$

Bridged-T Type:

$$\cosh \Gamma = 1 + \frac{2z_a z_b}{z_a(z_a + 4z_c) + 4z_b z_c} \tag{112}$$

and

$$K = \sqrt{\frac{z_a z_b (z_a + 4z_c)}{4(z_a + z_b)}}. \tag{113}$$

As an aid in obtaining the propagation constant, $\Gamma = A + iB$, from any of the three hyperbolic cosine formulæ it will be found convenient to use the following formulæ.

Computation Formulæ for the Complex Anti-Hyperbolic Cosine

It is known that many formulæ have already been derived for such evaluations but those below appear to give accurate results more readily.

Let it be desired to obtain A and B from the formula

$$\cosh (A + iB) = x + iy, \tag{114}$$

wherein x and y are known. A transformation of the x and y variables is first made so as to use the form of substitution and formulæ given in *B. S. T. J.*, October, 1924, pages 577 and 578. A further substitution and the application of hyperbolic formulæ give the following results where

$$\begin{aligned} U &= \tfrac{1}{2}(x - 1), \\ V &= \tfrac{1}{2}y, \\ P &= 4(U + U^2 + V^2), \end{aligned} \tag{115}$$

and

$$Q = \tfrac{1}{2} \sinh^{-1} \left| \frac{V}{U + U^2 + V^2} \right|.$$

When P is Positive:

$$A = \sinh^{-1} (\sqrt{P} \cosh Q)$$

and

$$B = \pm \sin^{-1} (\sqrt{P} \sinh Q). \tag{116}$$

When P is Negative:

$$A = \sinh^{-1} (\sqrt{-P} \sinh Q)$$

and

$$B = \pm \sin^{-1} (\sqrt{-P} \cosh Q). \tag{117}$$

When P is Zero, a Special Case:

$$A = \sinh^{-1} \sqrt{2|V|} = \tfrac{1}{2} \cosh^{-1} (1 + 4|V|)$$

and

$$B = \pm \sin^{-1} \sqrt{2|V|} = \pm \tfrac{1}{2} \cos^{-1} (1 - 4|V|). \tag{118}$$

In All Cases:

$$B = \cos^{-1}\left(\frac{1 + 2U}{\cosh A}\right) = \sin^{-1}\left(\frac{2V}{\sinh A}\right). \qquad (119)$$

The latter anti-cosine formula is particularly useful when B is in the neighborhood of $(2n + 1)\pi/2$, and both formulæ of (119) when considered together determine the sign of B.

The above formulæ give the solution of (114) which has a positive value for A (as in the propagation constant of a passive network). The other solution, since $\cosh(-\Gamma) = \cosh\Gamma$, would have values for both A and B which are the negative of those in the first solution (as may be possible in an active network).

It has been found that, when x and y are given to five or six decimals, it is possible to derive A and B to about this same degree of accuracy from these formulæ and the Smithsonian Mathematical Tables of Hyperbolic Functions. The formulæ may be used to advantage in accurately obtaining the propagation constant of a loaded line where x and y are calculated from the known circuit constants. (See footnote 2.)

Appendix IV

Propagation Characteristics and Formulæ for Various Lattice Type Networks

Networks of the lattice type only are specifically considered here since they have more general propagation characteristics than ladder or bridged-T types. However, transformations of any lattice type design obtained can be made to equivalent networks of these other types, if physical, by means of the simple relations given in Table II and the corresponding Section 2.5.

The network drawings show only half of the elements so as to avoid confusion; it is to be understood that the broken lines indicate the other series and lattice branches, respectively identical. The double subscript notation adopted for the elements is to be interpreted as follows: the first subscript on any element denotes the general position of the element in the network, 1 for the series branch and 2 for the lattice branch; the second subscript denotes the serial number of the element in either branch. Elements in the two branches which have the same serial numbers for their second subscripts correspond to each other according to the inverse network relations.

This group of networks, while not exhaustive, includes the simpler and perhaps most useful structures, but it could readily be extended. The propagation characteristics shown for each structure and derived

from computed results are representative and serve to give an idea of the possibilities of the network for design purposes. All networks except the last have a constant resistance iterative impedance R. Networks 1*a*–12 have attenuation so that they will usually be designed from their attenuation characteristics in terms of which the formulæ are given. There is usually more than one physical solution from the same attenuation characteristic, and in Networks 9 and 10 as many as four have been found possible. These multiple solutions all have different phase constants. A possible practical advantage of one solution over another may lie either in its phase constant or the magnitudes of its elements. It is of interest to point out that if these networks were designed from the phase characteristic some of them might have multiple solutions with different attenuation characteristics. For example, Networks 3*a* and 3*b* corresponding to the phase characteristics 1′ and 2′ each can have two such solutions.

The Networks 1*b*, 2*b*, etc., with their output terminals interchanged are, respectively, identical with Networks 1*a*, 2*a*, etc. Hence, any pair of these networks have the same attenuation constants but phase constants differing by π radians. An extension of this list to include Networks 6*b*, 7*b*, etc., was not thought to be necessary.

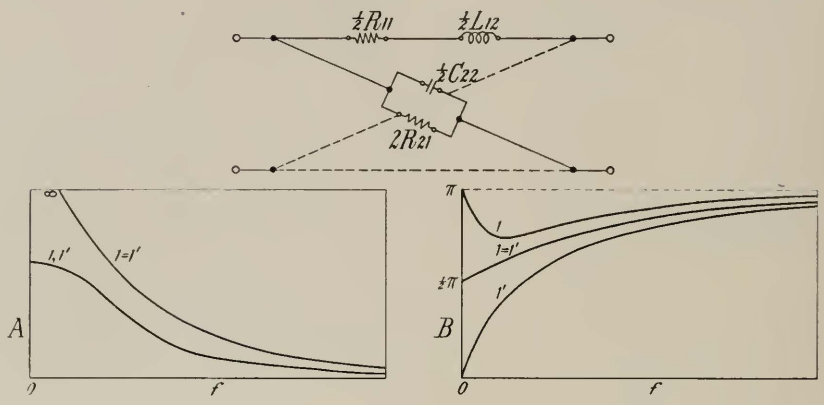
Several networks may have the same form of frequency function for F or H. Some values of the attenuation or phase coefficients will give a physical structure to one network but not to another. Whether a network having a definite A- or B-characteristic is physical or not can be determined most readily by a direct substitution of the coefficients in the formulæ for the elements. In certain cases these latter formulæ show easily that one network may give a physical result where another cannot. For example, Networks 6 and 10 both have the same F formula, but when one network is physical the other is not; similarly with Networks 7 and 9. These particular results would be expected from the fact that those pairs of networks cannot have the same attenuation characteristics, as seen from their structures.

Networks 13–17 have no attenuation and are designed from their phase characteristics. Network 18 represents a somewhat general form of artificial line and has other types of formulæ.

Examples of networks which are potentially complementary are Networks 1*a* and 2*b*; 1*b* and 2*a*; 3*a* and 3*b*; 11 and 12.

Transformations of impedance branches to equivalent ones can be made in some of the networks by means of the general transformation formulæ given in *B. S. T. J.*, January, 1923, pages 45 and 46.

NETWORK 1*a*



$$R_{11} = 2a_0R; \qquad L_{12} = \frac{a_1R}{\pi}.$$

$$R_{11}R_{21} = L_{12}/C_{22} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{P_0 + f^2}{Q_0 + f^2}.$$

Attenuation Linear Equation:

$$P_0 - FQ_0 = f^2(F - 1).$$

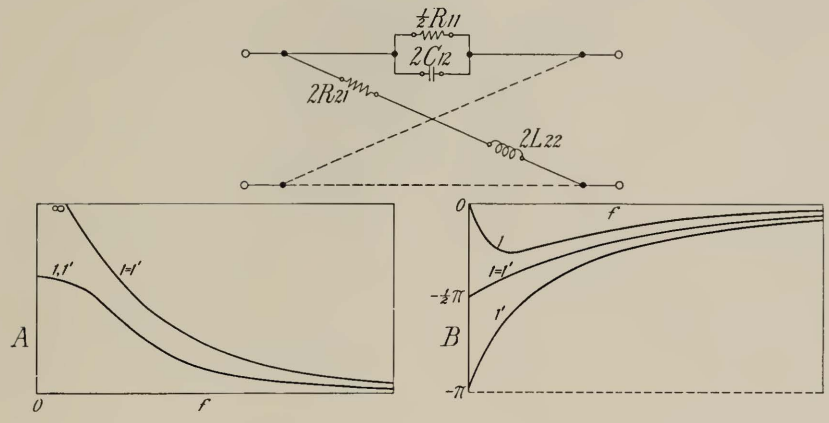
In physical solutions $O \leqq Q_0 \leqq P_0$.

$$1.\ a_0 = \frac{\sqrt{P_0} + \sqrt{Q_0}}{\sqrt{P_0} - \sqrt{Q_0}}; \qquad a_1 = \frac{2}{\sqrt{P_0} - \sqrt{Q_0}}.$$

$$1'.\ a_0 = \frac{\sqrt{P_0} - \sqrt{Q_0}}{\sqrt{P_0} + \sqrt{Q_0}}; \qquad a_1 = \frac{2}{\sqrt{P_0} + \sqrt{Q_0}}.$$

$$II = \tan B = \frac{2a_1f}{(1 - a_0^2) - a_1^2f^2}.$$

NETWORK 1*b*



$$R_{11} = 2a_0R; \qquad C_{12} = \frac{b_1}{4\pi a_0 R}.$$

$$R_{11}R_{21} = L_{22}/C_{12} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{P_0 + f^2}{Q_0 + f^2}.$$

Attenuation Linear Equation:

$$P_0 - FQ_0 = f^2(F - 1).$$

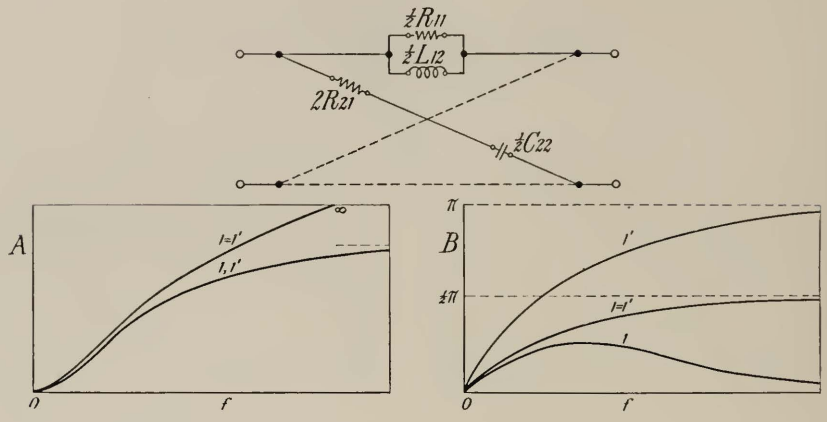
In physical solutions $O \leqq Q_0 \leqq P_0$.

$$1.\ a_0 = \frac{\sqrt{P_0} - \sqrt{Q_0}}{\sqrt{P_0} + \sqrt{Q_0}}; \qquad b_1 = \frac{2}{\sqrt{P_0} + \sqrt{Q_0}}.$$

$$1'.\ a_0 = \frac{\sqrt{P_0} + \sqrt{Q_0}}{\sqrt{P_0} - \sqrt{Q_0}}; \qquad b_1 = \frac{2}{\sqrt{P_0} - \sqrt{Q_0}}.$$

$$II = \tan B = \frac{-2a_0b_1f}{(1 - a_0^2) + b_1^2f^2}.$$

NETWORK 2*a*



$$R_{11} = \frac{2a_1R}{b_1}; \qquad L_{12} = \frac{a_1R}{\pi}.$$

$$R_{11}R_{21} = L_{12}/C_{22} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{1 + P_2f^2}{1 + Q_2f^2}.$$

Attenuation Linear Equation:

$$P_2 - FQ_2 = (F - 1)/f^2.$$

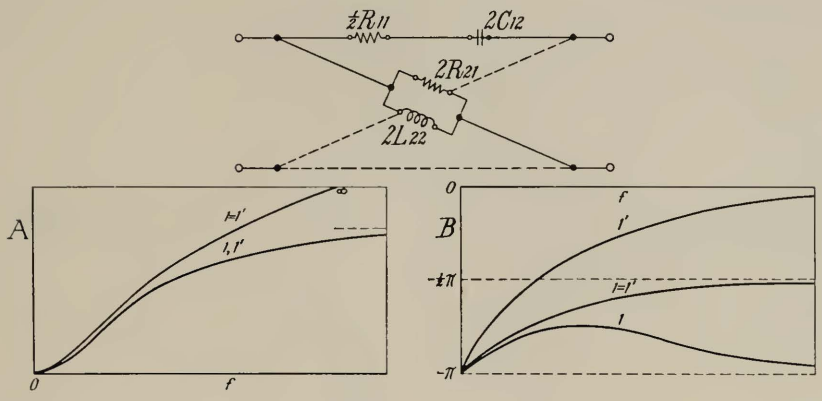
In physical solutions $O \leqq Q_2 \leqq P_2$.

$$1.\ \left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} = \tfrac{1}{2}(\sqrt{P_2} \mp \sqrt{Q_2}).$$

$$1'.\ \left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} = \tfrac{1}{2}(\sqrt{P_2} \pm \sqrt{Q_2}).$$

$$H = \tan B = \frac{2a_1f}{1 - (a_1^2 - b_1^2)f^2}.$$

NETWORK 2*b*



$$R_{11} = \frac{2a_1R}{b_1}; \qquad C_{12} = \frac{b_1}{4\pi R}.$$

$$R_{11}R_{21} = L_{22}/C_{12} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{1 + P_2f^2}{1 + Q_2f^2}.$$

Attenuation Linear Equation:

$$P_2 - FQ_2 = (F - 1)/f^2.$$

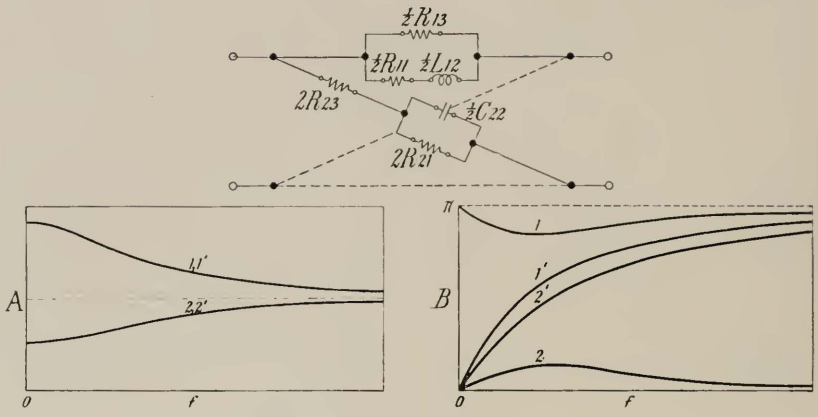
In physical solutions $O \leqq Q_2 \leqq P_2$.

$$1.\ \left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} = \tfrac{1}{2}(\sqrt{P_2} \pm \sqrt{Q_2}).$$

$$1'.\ \left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} = \tfrac{1}{2}(\sqrt{P_2} \mp \sqrt{Q_2}).$$

$$H = \tan B = \frac{2b_1f}{1 + (a_1^2 - b_1^2)f^2}.$$

NETWORK 3a



$$R_{11} = \frac{2a_0a_1R}{a_1b_0 - a_0}; \qquad L_{12} = \frac{a_1^2R}{\pi(a_1b_0 - a_0)}; \qquad R_{13} = 2a_1R.$$

$$R_{11}R_{21} = L_{12}/C_{22} = R_{13}R_{23} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{P_0 + f^2}{Q_0 + Q_2f^2}.$$

Attenuation Linear Equation:

$$-P_0 + FQ_0 + f^2FQ_2 = f^2.$$

In physical solutions $O \leqq Q_0 \leqq P_0$; $O \leqq Q_2 \leqq 1$.

If $Q_0 < P_0Q_2$ (A decreases with frequency):

$$1.\ \left.\begin{matrix} a_0 \\ b_0 \end{matrix}\right\} = \frac{\sqrt{P_0} \pm \sqrt{Q_0}}{1 - \sqrt{Q_2}}; \qquad a_1 = \frac{1 + \sqrt{Q_2}}{1 - \sqrt{Q_2}}.$$

$$1'.\ \left.\begin{matrix} a_0 \\ b_0 \end{matrix}\right\} = \frac{\sqrt{P_0} \mp \sqrt{Q_0}}{1 - \sqrt{Q_2}}; \qquad a_1 = \frac{1 + \sqrt{Q_2}}{1 - \sqrt{Q_2}}.$$

If $Q_0 > P_0Q_2$ (A increases with frequency):

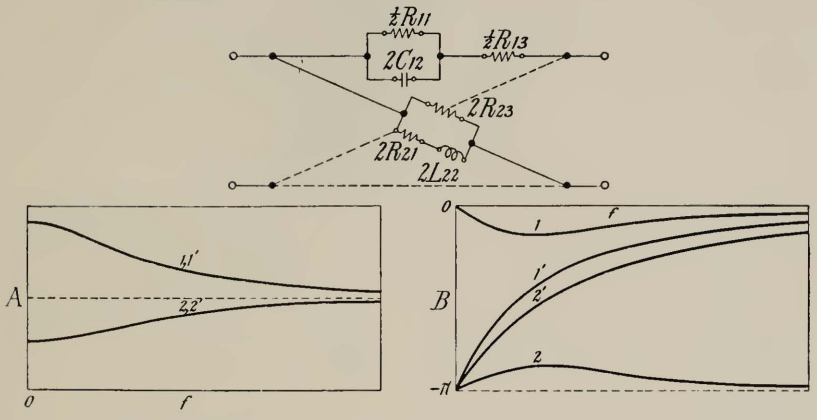
$$2.\ \left.\begin{matrix} a_0 \\ b_0 \end{matrix}\right\} = \frac{\sqrt{P_0} \mp \sqrt{Q_0}}{1 + \sqrt{Q_2}}; \qquad a_1 = \frac{1 - \sqrt{Q_2}}{1 + \sqrt{Q_2}}.$$

2'. Same formulæ as in 1'.

If $Q_0 = P_0Q_2$, $F = 1/Q_2$ (A is constant).

$$II = \tan B = \frac{2(a_1b_0 - a_0)f}{(b_0^2 - a_0^2) + (1 - a_1^2)f^2}.$$

NETWORK 3*b*



$$R_{11} = \frac{2(a_0b_1 - a_1)R}{b_1}; \qquad C_{12} = \frac{b_1^2}{4\pi(a_0b_1 - a_1)R}; \qquad R_{13} = \frac{2a_1R}{b_1}.$$

$$R_{11}R_{21} = L_{22}/C_{12} = R_{13}R_{23} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{P_0 + f^2}{Q_0 + Q_2f^2}.$$

Attenuation Linear Equation:

$$-P_0 + FQ_0 + f^2FQ_2 = f^2.$$

In physical solutions $O \leqq Q_0 \leqq P_0$; $O \leqq Q_2 \leqq 1$.

If $Q_0 < P_0Q_2$ (A decreases with frequency):

$$1.\ a_0 = \frac{\sqrt{P_0} - \sqrt{Q_0}}{\sqrt{P_0} + \sqrt{Q_0}}; \qquad \left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} = \frac{1 \mp \sqrt{Q_2}}{\sqrt{P_0} + \sqrt{Q_0}}.$$

$$1'.\ a_0 = \frac{\sqrt{P_0} + \sqrt{Q_0}}{\sqrt{P_0} - \sqrt{Q_0}}; \qquad \left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} = \frac{1 \mp \sqrt{Q_2}}{\sqrt{P_0} - \sqrt{Q_0}}.$$

If $Q_0 > P_0Q_2$ (A increases with frequency):

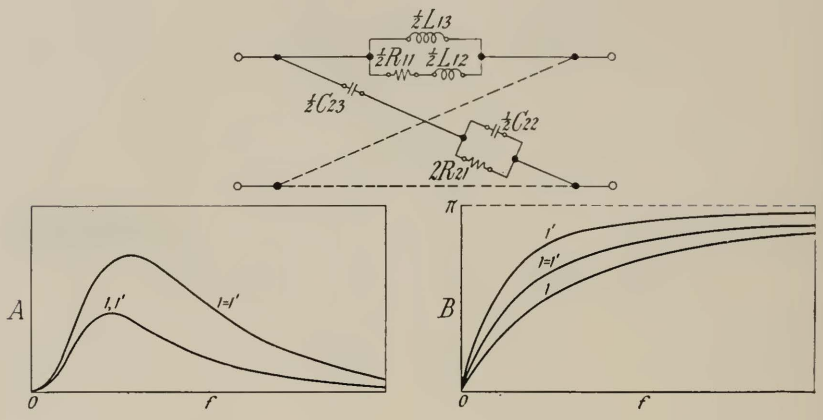
$$2.\ a_0 = \frac{\sqrt{P_0} + \sqrt{Q_0}}{\sqrt{P_0} - \sqrt{Q_0}}; \qquad \left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} = \frac{1 \pm \sqrt{Q_2}}{\sqrt{P_0} - \sqrt{Q_0}}.$$

2′. Same formulæ as in 1′.

If $Q_0 = P_0Q_2$, $F = 1/Q_2$ (A is constant).

$$H = \tan B = \frac{-2(a_0b_1 - a_1)f}{(1 - a_0^2) + (b_1^2 - a_1^2)f^2}.$$

NETWORK 4*a*



$$R_{11} = \frac{2a_1^2R}{a_1b_1 - a_2}; \qquad L_{12} = \frac{a_1a_2R}{\pi(a_1b_1 - a_2)}; \qquad L_{13} = \frac{a_1R}{\pi}.$$

$$R_{11}R_{21} = L_{12}/C_{22} = L_{13}/C_{23} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{1 + P_2f^2 + P_4f^4}{1 + Q_2f^2 + P_4f^4}.$$

Attenuation Linear Equation:

$$P_2 - f^2(F - 1)P_4 - FQ_2 = (F - 1)/f^2.$$

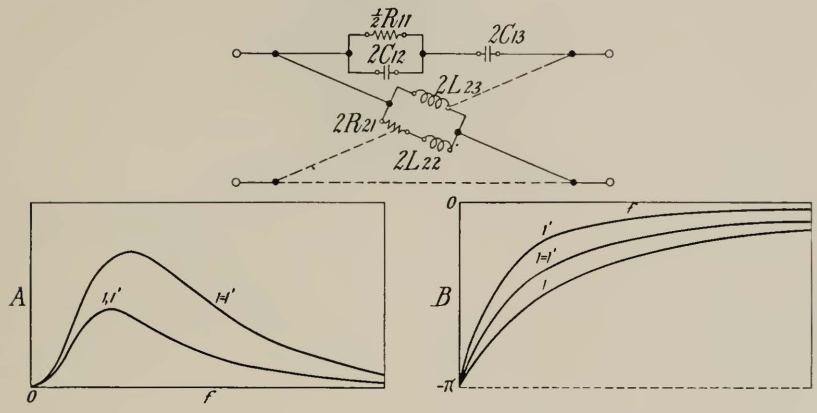
In physical solutions $O \leqq 2\sqrt{P_4} \leqq Q_2 \leqq P_2$.

1. $\left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} = \frac{1}{2}(\sqrt{P_2 + 2\sqrt{P_4}} \mp \sqrt{Q_2 - 2\sqrt{P_4}}); \qquad a_2 = \sqrt{P_4}.$

1′. Same formulæ as in 1, but with a_1 and b_1 interchanged.

$$H = \tan B = \frac{2a_1f + 2a_2b_1f^3}{1 - (a_1^2 - b_1^2)f^2 - a_2^2f^4}.$$

NETWORK 4*b*



$$R_{11} = \frac{2(a_1b_1 - b_2)R}{b_1^2}; \qquad C_{12} = \frac{b_1b_2}{4\pi(a_1b_1 - b_2)R}; \qquad C_{13} = \frac{b_1}{4\pi R}.$$

$$R_{11}R_{21} = L_{22}/C_{12} = L_{23}/C_{13} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{1 + P_2f^2 + P_4f^4}{1 + Q_2f^2 + P_4f^4}.$$

Attenuation Linear Equation:

$$P_2 - f^2(F - 1)P_4 - FQ_2 = (F - 1)/f^2.$$

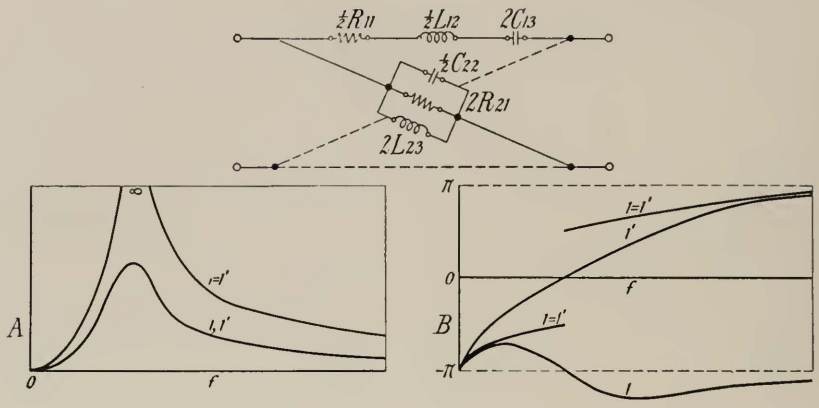
In physical solutions $O \leqq 2\sqrt{P_4} \leqq Q_2 \leqq P_2$.

1. $\left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} = \frac{1}{2}(\sqrt{P_2 + 2\sqrt{P_4}} \pm \sqrt{Q_2 - 2\sqrt{P_4}}); \qquad b_2 = \sqrt{P_4}.$

1′. Same formulæ as in 1, but with a_1 and b_1 interchanged.

$$H = \tan B = \frac{2b_1f + 2a_1b_2f^3}{1 + (a_1^2 - b_1^2)f^2 - b_2^2f^4}.$$

NETWORK 5*a*



$$R_{11} = \frac{2a_1R}{b_1}; \qquad L_{12} = \frac{a_2R}{\pi b_1}; \qquad C_{13} = \frac{b_1}{4\pi R}.$$

$$R_{11}R_{21} = L_{12}/C_{22} = L_{23}/C_{13} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{1 + P_2f^2 + P_4f^4}{1 + Q_2f^2 + P_4f^4}.$$

Attenuation Linear Equation:

$$P_2 - FQ_2 - f^2(F - 1)P_4 = (F - 1)/f^2.$$

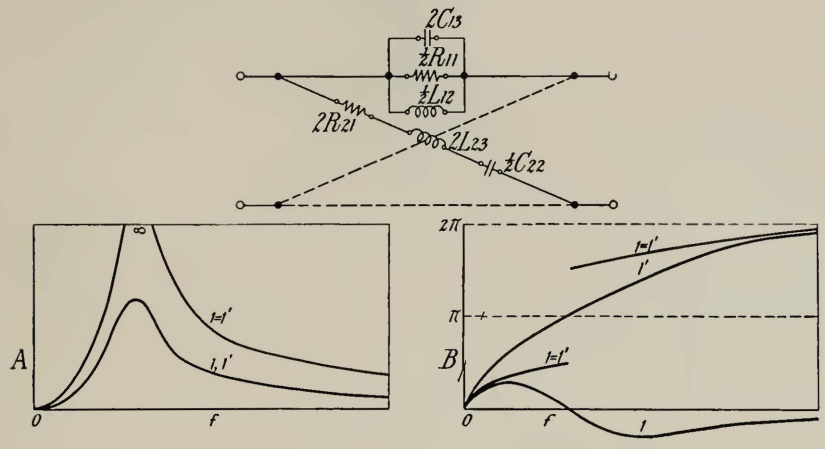
In physical solutions $-2\sqrt{P_4} \leqq Q_2 \leqq P_2$.

$$1.\ \left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} = \tfrac{1}{2}(\sqrt{P_2 + 2\sqrt{P_4}} \pm \sqrt{Q_2 + 2\sqrt{P_4}}); \qquad a_2 = \sqrt{P_4}.$$

1′. Same formulæ as in 1, but with a_1 and b_1 interchanged.

$$II = \tan B = \frac{2b_1f - 2a_2b_1f^3}{1 + (a_1^2 - 2a_2 - b_1^2)f^2 + a_2^2f^4}.$$

NETWORK 5*b*



$$R_{11} = \frac{2a_1R}{b_1}; \qquad L_{12} = \frac{a_1R}{\pi}; \qquad C_{13} = \frac{b_2}{4\pi a_1 R}.$$

$$R_{11}R_{21} = L_{12}/C_{22} = L_{23}/C_{13} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{1 + P_2f^2 + P_4f^4}{1 + Q_2f^2 + P_4f^4}.$$

Attenuation Linear Equation:

$$P_2 - FQ_2 - f^2(F - 1)P_4 = (F - 1)/f^2.$$

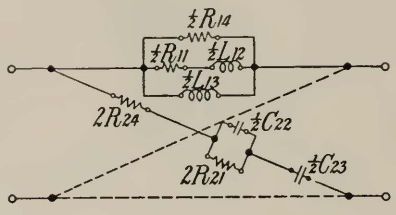
In physical solutions $-2\sqrt{P_4} \leqq Q_2 \leqq P_2$.

$$1.\ \left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} = \tfrac{1}{2}\left(\sqrt{P_2 + 2\sqrt{P_4}} \mp \sqrt{Q_2 + 2\sqrt{P_4}}\right); \qquad b_2 = \sqrt{P_4}.$$

1′. Same formulæ as in 1, but with a_1 and b_1 interchanged.

$$H = \tan B = \frac{2a_1f - 2a_1b_2f^3}{1 - (a_1^2 - b_1^2 + 2b_2)f^2 + b_2^2f^4}.$$

NETWORK 6



A- and B-characteristics are similar to those of Networks 2a and 4a.

$$R_{11} = \frac{2a_1^2a_2R}{a_1a_2b_1 - a_1^2b_2 - a_2^2}; \qquad L_{12} = \frac{a_1a_2^2R}{\pi(a_1a_2b_1 - a_1^2b_2 - a_2^2)};$$

$$L_{13} = \frac{a_1R}{\pi}; \qquad R_{14} = \frac{2a_2R}{b_2}.$$

$$R_{11}R_{21} = L_{12}/C_{22} = L_{13}/C_{23} = R_{14}R_{24} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{1 + P_2f^2 + P_4f^4}{1 + Q_2f^2 + Q_4f^4}.$$

Attenuation Linear Equation:

$$P_2 + f^2P_4 - FQ_2 - f^2FQ_4 = (F - 1)/f^2.$$

In unrestricted solutions, where $O \leqq Q_4 \leqq P_4$:

$$\left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} \text{ or } \left.\begin{matrix} b_1 \\ a_1 \end{matrix}\right\} = \tfrac{1}{2}(\sqrt{P_2 + 2\sqrt{P_4}} \pm \sqrt{Q_2 - 2\sqrt{Q_4}});$$

$$\left.\begin{matrix} a_2 \\ b_2 \end{matrix}\right\} = \tfrac{1}{2}(\sqrt{P_4} \pm \sqrt{Q_4}).$$

Also

$$\left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} \text{ or } \left.\begin{matrix} b_1 \\ a_1 \end{matrix}\right\} = \tfrac{1}{2}(\sqrt{P_2 + 2\sqrt{P_4}} \pm \sqrt{Q_2 + 2\sqrt{Q_4}});$$

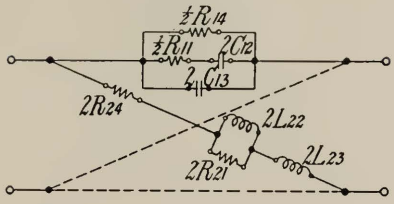
$$\left.\begin{matrix} a_2 \\ b_2 \end{matrix}\right\} = \tfrac{1}{2}(\sqrt{P_4} \mp \sqrt{Q_4}).$$

In physical solutions a_1, a_2, b_1, b_2 are positive;

$$a_1^2b_2 + a_2^2 \leqq a_1a_2b_1.$$

$$II = \tan B = \frac{2a_1f - 2(a_1b_2 - a_2b_1)f^3}{1 - (a_1^2 - b_1^2 + 2b_2)f^2 - (a_2^2 - b_2^2)f^4}.$$

NETWORK 7



A- and B-characteristics are similar to those of Networks 1b and 4b.

$$R_{11} = \frac{2a_0a_1^2R}{a_0a_1b_1 - a_1^2 - a_0^2b_2}; \qquad C_{12} = \frac{a_0a_1b_1 - a_1^2 - a_0^2b_2}{4\pi a_0^2a_1R};$$

$$C_{13} = \frac{b_2}{4\pi a_1R}; \qquad R_{14} = 2a_0R.$$

$$R_{11}R_{21} = L_{22}/C_{12} = L_{23}/C_{13} = R_{14}R_{24} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{P_0 + P_2f^2 + f^4}{Q_0 + Q_2f^2 + f^4}.$$

Attenuation Linear Equation:

$$P_0 + f^2P_2 - FQ_0 - f^2FQ_2 = f^4(F - 1).$$

In unrestricted solutions, where $O \leqq Q_0 \leqq P_0$:

$$a_0 = \frac{\sqrt{P_0} - \sqrt{Q_0}}{\sqrt{P_0} + \sqrt{Q_0}}; \qquad b_2 = \frac{2}{\sqrt{P_0} + \sqrt{Q_0}};$$

$$\left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} \text{or} \left.\begin{matrix} b_1 \\ a_1 \end{matrix}\right\} = \frac{\sqrt{P_2 + 2\sqrt{P_0}} \pm \sqrt{Q_2 + 2\sqrt{Q_0}}}{\sqrt{P_0} + \sqrt{Q_0}}.$$

Also

$$a_0 = \frac{\sqrt{P_0} + \sqrt{Q_0}}{\sqrt{P_0} - \sqrt{Q_0}}; \qquad b_2 = \frac{2}{\sqrt{P_0} - \sqrt{Q_0}};$$

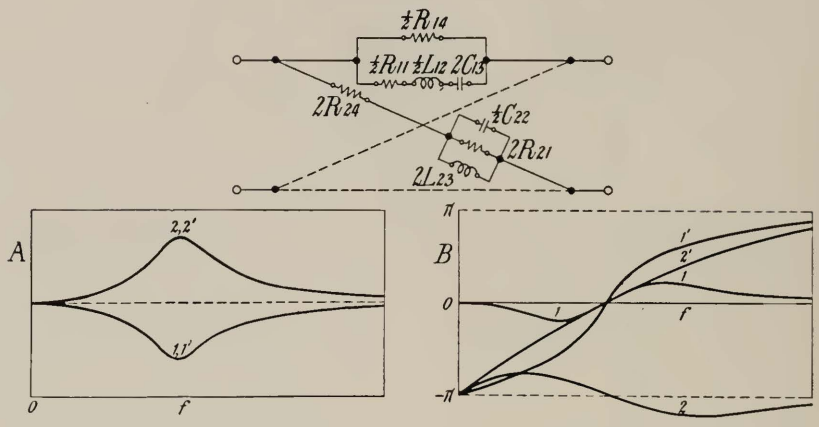
$$\left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} \text{or} \left.\begin{matrix} b_1 \\ a_1 \end{matrix}\right\} = \frac{\sqrt{P_2 + 2\sqrt{P_0}} \pm \sqrt{Q_2 - 2\sqrt{Q_0}}}{\sqrt{P_0} - \sqrt{Q_0}}.$$

In physical solutions a_0, a_1, b_1, b_2 are positive;

$$a_1^2 + a_0^2b_2 \leqq a_0a_1b_1.$$

$$H = \tan B = \frac{2(a_1 - a_0b_1)f - 2a_1b_2f^3}{(1 - a_0^2) - (a_1^2 - b_1^2 + 2b_2)f^2 + b_2^2f^4}.$$

NETWORK 8



$$R_{11} = \frac{2a_0a_1R}{a_0b_1 - a_1}; \qquad L_{12} = \frac{a_0^2b_2R}{\pi(a_0b_1 - a_1)};$$

$$C_{13} = \frac{a_0b_1 - a_1}{4\pi a_0^2R}; \qquad R_{14} = 2a_0R.$$

$$R_{11}R_{21} = L_{12}/C_{22} = L_{23}/C_{13} = R_{14}R_{24} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{F_0 + P_2f^2 + F_0Q_4f^4}{1 + Q_2f^2 + Q_4f^4}.$$

$$F_0(f = o) = F_\infty(f = \infty); \qquad A_0 = A_\infty.$$

Attenuation Linear Equation:

$$-P_2 + FQ_2 - f^2(F_0 - F)Q_4 = (F_0 - F)/f^2.$$

In physical solutions $O \leqq Q_2 + 2\sqrt{Q_4} \equiv n \leqq P_2 + 2F_0\sqrt{Q_4} \equiv m$.
If $P_2 < F_0Q_2$ (A has a minimum):

1. $a_0 = \tanh(A_0/2); \qquad b_2 = \sqrt{Q_4};$

$$\left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} = \tfrac{1}{2}(1 - \tanh(A_0/2))(\sqrt{m} \mp \sqrt{n}).$$

1'. $a_0 = \coth(A_0/2); \qquad b_2 = \sqrt{Q_4};$

$$\left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} = \tfrac{1}{2}(\coth(A_0/2) - 1)(\sqrt{m} \mp \sqrt{n}).$$

If $P_2 > F_0Q_2$ (A has a maximum):

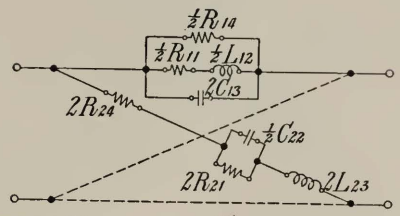
2. $a_0 = \coth(A_0/2); \qquad b_2 = \sqrt{Q_4};$

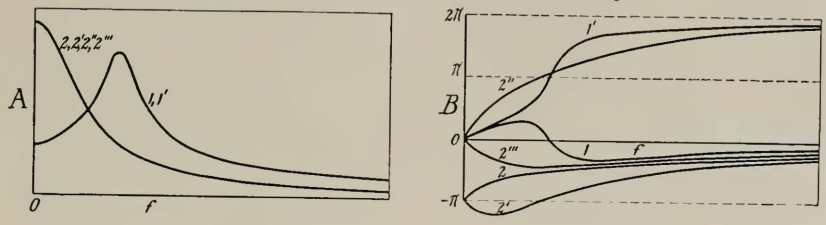
$$\left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} = \tfrac{1}{2}(\coth(A_0/2) - 1)(\sqrt{m} \pm \sqrt{n}).$$

2'. The same formulæ as in 1'.

$$H = \tan B = \frac{2(a_0b_1 - a_1)(-f + b_2f^3)}{(1 - a_0^2) - (a_1^2 - b_1^2 + 2(1 - a_0^2)b_2)f^2 + (1 - a_0^2)b_2^2f^4}.$$

NETWORK 9





$$R_{11} = \frac{2a_0a_1^2R}{a_0^2b_2 + a_1^2 - a_0a_1b_1}; \qquad L_{12} = \frac{a_1^3R}{\pi(a_0^2b_2 + a_1^2 - a_0a_1b_1)};$$

$$C_{13} = \frac{b_2}{4\pi a_1R}; \qquad R_{14} = \frac{2a_1^2R}{a_1b_1 - a_0b_2}.$$

$$R_{11}R_{21} = L_{12}/C_{22} = L_{23}/C_{13} = R_{14}R_{24} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{P_0 + P_2f^2 + f^4}{Q_0 + Q_2f^2 + f^4}.$$

Attenuation Linear Equation:

$$P_0 + f^2P_2 - FQ_0 - f^2FQ_2 = f^4(F - 1).$$

In unrestricted solutions, where $O \leqq Q_0 \leqq P_0$:

$$a_0 = \frac{\sqrt{P_0} - \sqrt{Q_0}}{\sqrt{P_0} + \sqrt{Q_0}}; \qquad b_2 = \frac{2}{\sqrt{P_0} + \sqrt{Q_0}};$$

$$\left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} \text{or} \left.\begin{matrix} b_1 \\ a_1 \end{matrix}\right\} = \frac{\sqrt{P_2 + 2\sqrt{P_0}} \pm \sqrt{Q_2 + 2\sqrt{Q_0}}}{\sqrt{P_0} + \sqrt{Q_0}}.$$

Also

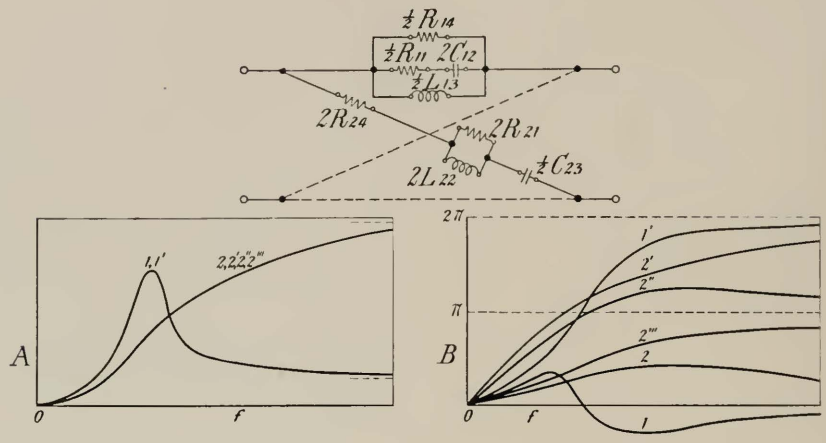
$$a_0 = \frac{\sqrt{P_0} + \sqrt{Q_0}}{\sqrt{P_0} - \sqrt{Q_0}}; \qquad b_2 = \frac{2}{\sqrt{P_0} - \sqrt{Q_0}};$$

$$\left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} \text{or} \left.\begin{matrix} b_1 \\ a_1 \end{matrix}\right\} = \frac{\sqrt{P_2 + 2\sqrt{P_0}} \pm \sqrt{Q_2 - 2\sqrt{Q_0}}}{\sqrt{P_0} - \sqrt{Q_0}}.$$

In physical solutions $Q_2 \leqq P_2$; $a_0a_1b_1 \leqq a_0^2b_2 + a_1^2$.

$$H = \tan B = \frac{2(a_1 - a_0b_1)f - 2a_1b_2f^3}{(1 - a_0^2) - (a_1^2 - b_1^2 + 2b_2)f^2 + b_2^2f^4}.$$

NETWORK 10



$$R_{11} = \frac{2a_1^2a_2R}{a_1^2b_2 + a_2^2 - a_1a_2b_1}; \qquad C_{12} = \frac{a_1^2b_2 + a_2^2 - a_1a_2b_1}{4\pi a_1^3R};$$

$$L_{13} = \frac{a_1R}{\pi}; \qquad R_{14} = \frac{2a_1^2R}{a_1b_1 - a_2}.$$

$$R_{11}R_{21} = L_{22}/C_{12} = L_{13}/C_{23} = R_{14}R_{24} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{1 + P_2f^2 + P_4f^4}{1 + Q_2f^2 + Q_4f^4}.$$

Attenuation Linear Equation:

$$P_2 + f^2P_4 - FQ_2 - f^2FQ_4 = (F - 1)/f^2.$$

In unrestricted solutions, where $O \leqq Q_4 \leqq P_4$:

$$\left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} \text{ or } \left.\begin{matrix} b_1 \\ a_1 \end{matrix}\right\} = \tfrac{1}{2}(\sqrt{P_2 + 2\sqrt{P_4}} \pm \sqrt{Q_2 - 2\sqrt{Q_4}});$$

$$\left.\begin{matrix} a_2 \\ b_2 \end{matrix}\right\} = \tfrac{1}{2}(\sqrt{P_4} \pm \sqrt{Q_4}).$$

Also

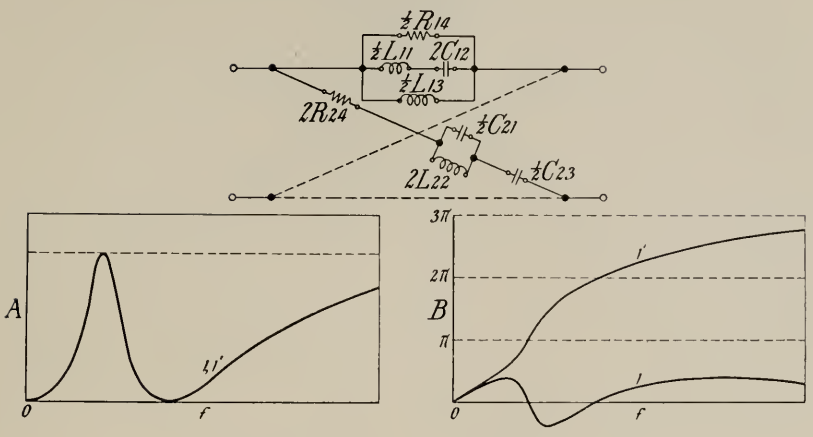
$$\left.\begin{matrix} a_1 \\ b_1 \end{matrix}\right\} \text{ or } \left.\begin{matrix} b_1 \\ a_1 \end{matrix}\right\} = \tfrac{1}{2}(\sqrt{P_2 + 2\sqrt{P_4}} \pm \sqrt{Q_2 + 2\sqrt{Q_4}});$$

$$\left.\begin{matrix} a_2 \\ b_2 \end{matrix}\right\} = \tfrac{1}{2}(\sqrt{P_4} \mp \sqrt{Q_4}).$$

In physical solutions $Q_2 \leqq P_2$; $a_1a_2b_1 \leqq a_1^2b_2 + a_2^2$.

$$II = \tan B = \frac{2a_1f - 2(a_1b_2 - a_2b_1)f^3}{1 - (a_1^2 - b_1^2 + 2b_2)f^2 - (a_2^2 - b_2^2)f^4}$$

NETWORK 11



$$L_{11} = \frac{a_1 a_3 R}{\pi(a_1 b_2 - a_3)}; \qquad C_{12} = \frac{a_1 b_2 - a_3}{4\pi a_1^2 R};$$

$$L_{13} = \frac{a_1 R}{\pi}; \qquad R_{14} = \frac{2R}{m}.$$

$$L_{11}/C_{21} = L_{22}/C_{12} = L_{13}/C_{23} = R_{14}R_{24} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{1 + (1+m)^2 y^2}{1 + (1-m)^2 y^2},$$

where

$$y = \frac{a_1 f - a_3 f^3}{1 - b_2 f^2}$$

is the total parallel reactance in z_{11} divided by $2R$.

1. $m = \coth \frac{1}{2} A_\infty$;

1'. $m = \tanh \frac{1}{2} A_\infty$;

where A_∞ is the maximum attenuation at $f = \infty$ and at the internal frequency $f = 1/\sqrt{b_2}$.

Attenuation Linear Equation:

$$a_1 - f^2 a_3 + f y b_2 = y/f,$$

where

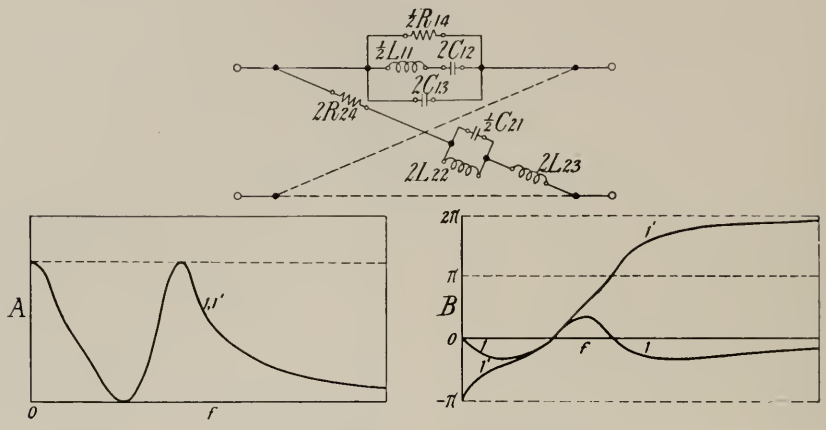
$$y = \pm \sqrt{\frac{F-1}{(1+m)^2 - (1-m)^2 F}}$$

and the signs to be taken for y correspond to the particular reactance branches involved, whose signs in order on the frequency scale are $+$, $-$, and $+$.

In physical solutions $a_3 \leqq a_1 b_2$.

$$H = \tan B = \frac{2y}{1 - (1-m^2)y^2}.$$

NETWORK 12



$$L_{11} = \frac{a_2^2 R}{\pi(a_2 b_1 - b_3)}; \qquad C_{12} = \frac{a_2 b_1 - b_3}{4\pi a_2 R};$$

$$C_{13} = \frac{b_3}{4\pi a_2 R}; \qquad R_{14} = \frac{2R}{m}.$$

$$L_{11}/C_{21} = L_{22}/C_{12} = L_{23}/C_{13} = R_{14}R_{24} = R^2.$$

$$F = e^{2A} = 10^{\frac{TU}{10}} = \frac{1 + (1 + m)^2 y^2}{1 + (1 - m)^2 y^2},$$

where

$$y = \frac{-1 + a_2 f^2}{b_1 f - b_3 f^3}$$

is the total parallel reactance in z_{11} divided by $2R$.

1. $m = \coth \frac{1}{2} A_0$;

1'. $m = \tanh \frac{1}{2} A_0$;

where A_0 is the maximum attenuation at $f = 0$ and at the internal frequency $f = \sqrt{b_1/b_3}$.

Attenuation Linear Equation:

$$a_2 - (y/f)b_1 + fyb_3 = 1/f^2,$$

where

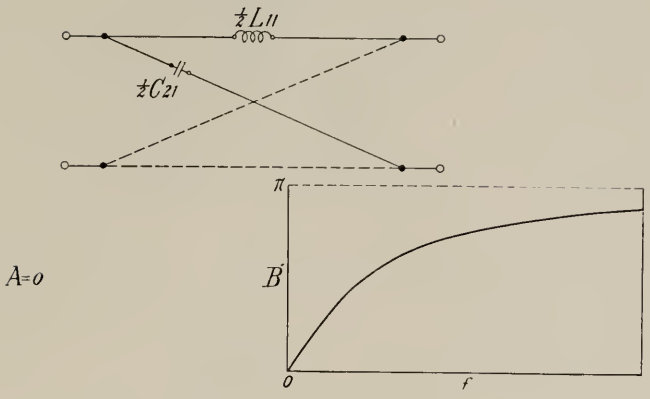
$$y = \pm \sqrt{\frac{F - 1}{(1 + m)^2 - (1 - m)^2 F}}$$

and the signs to be taken for y correspond to the particular reactance branches involved, whose signs in order on the frequency scale are $-$, $+$, and $-$.

In physical solutions $b_3 \leqq a_2 b_1$.

$$II = \tan B = \frac{2y}{1 - (1 - m^2)y^2}.$$

NETWORK 13



$$L_{11} = \frac{a_1 R}{\pi}; \qquad L_{11}/C_{21} = R^2.$$

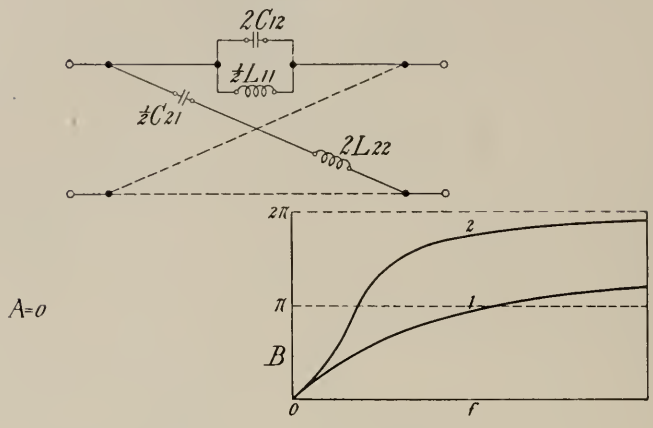
$$H = \tan \tfrac{1}{2}B = a_1 f.$$

Phase Linear Equation:

$$a_1 = H/f.$$

(See also formula (75).)

NETWORK 14



$$L_{11} = \frac{a_1 R}{\pi}; \qquad C_{12} = \frac{b_2}{4\pi a_1 R}.$$

$$L_{11}/C_{21} = L_{22}/C_{12} = R^2.$$

$$H = \tan \tfrac{1}{2}B = \frac{a_1 f}{1 - b_2 f^2}.$$

Phase Linear Equation:

$$a_1 + fHb_2 = H/f.$$

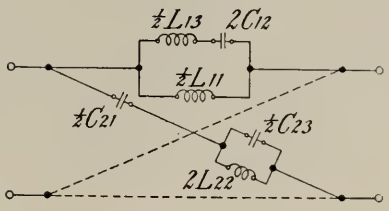
1. $b_2 < \frac{1}{3}a_1^2$. 2. $b_2 > \frac{1}{3}a_1^2$.

Equivalent Network, if $b_2 \leqq \frac{1}{4}a_1^2$:

Two sections (a_1' and a_1'') of Network 13:

$$\left.\begin{matrix} a_1' \\ a_1'' \end{matrix}\right\} = \tfrac{1}{2}(a_1 \pm \sqrt{a_1^2 - 4b_2}).$$

NETWORK 15



$A = 0$.

B-characteristic is the sum of those for Networks 13 and 14.

$$L_{11} = \frac{a_1 R}{\pi}; \qquad C_{12} = \frac{a_1 b_2 - a_3}{4\pi a_1^2 R}; \qquad L_{13} = \frac{a_1 a_3 R}{\pi(a_1 b_2 - a_3)}.$$

$$L_{11}/C_{21} = L_{22}/C_{12} = L_{13}/C_{23} = R^2.$$

$$H = \tan \tfrac{1}{2}B = \frac{a_1 f - a_3 f^3}{1 - b_2 f^2}.$$

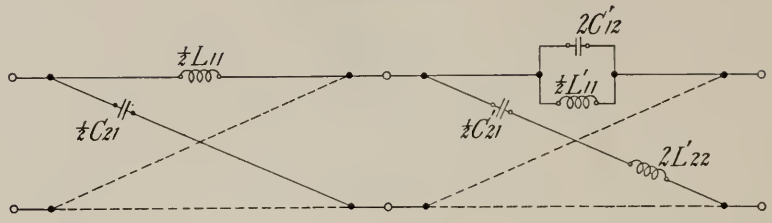
Phase Linear Equation:

$$a_1 - f^2 a_3 + fHb_2 = H/f.$$

In physical solutions $a_3 \leqq a_1 b_2$.

Equivalent to Network 16.

NETWORK 16



$A = 0$.

B-characteristic is the sum of those for Networks 13 and 14.

$$L_{11} = \frac{a_1 R}{\pi}; \qquad L_{11}' = \frac{a_1' R}{\pi}; \qquad C_{12}' = \frac{b_2'}{4\pi a_1' R}.$$

$$L_{11}/C_{21} = L_{11}'/C_{21}' = L_{22}'/C_{12}' = R^2.$$

$$H = \tan \tfrac{1}{2}B = \frac{M_1 f + M_3 f^3}{1 + N_2 f^2}.$$

Phase Linear Equation:

$$M_1 + f^2 M_3 - fHN_2 = H/f.$$

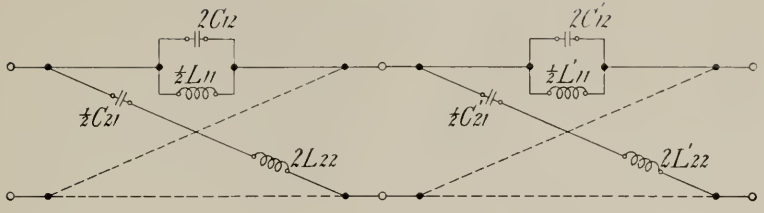
$$a_1^3 - M_1 a_1^2 - N_2 a_1 + M_3 = 0;$$

$$a_1' = M_1 - a_1;$$

$$b_2' = -M_3/a_1.$$

Equivalent to Network 15.

NETWORK 17



$A = 0$.

B-characteristic is the sum of those for two sections of Network 14.

$$L_{11} = \frac{a_1 R}{\pi}; \qquad C_{12} = \frac{b_2}{4\pi a_1 R}; \qquad L_{11}' = \frac{a_1' R}{\pi}; \qquad C_{12}' = \frac{b_2'}{4\pi a_1' R}.$$

$$L_{11}/C_{21} = L_{22}/C_{12} = L_{11}'/C_{21}' = L_{22}'/C_{12}' = R^2.$$

$$H = \tan \tfrac{1}{2}B = \frac{M_1 f + M_3 f^3}{1 + N_2 f^2 + N_4 f^4}.$$

Phase Linear Equation:

$$M_1 + f^2 M_3 - fHN_2 - f^3 HN_4 = H/f.$$

In physical solutions M_1 and N_4 are positive, as are also a_1, a_1', b_2, and b_2'. M_3 and N_2 are negative.

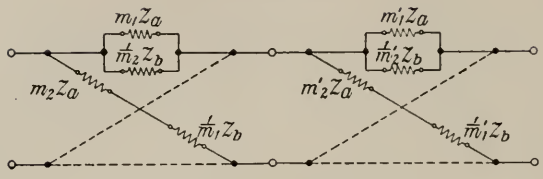
$$q^3 + 2N_2 q^2 + (-M_1 M_3 + N_2^2 - 4N_4)q + (M_1^2 N_4 - M_1 M_3 N_2 + M_3^2) = 0;$$

$$\left.\begin{matrix} a_1 \\ a_1' \end{matrix}\right\} = \tfrac{1}{2}(M_1 \pm \sqrt{M_1^2 - 4q});$$

$$\left.\begin{matrix} b_2 \\ b_2' \end{matrix}\right\} \text{ or } \left.\begin{matrix} b_2' \\ b_2 \end{matrix}\right\} = \tfrac{1}{2}(-(N_2 + q) \pm \sqrt{(N_2 + q)^2 - 4N_4}),$$

the determining condition being that $a_1 b_2' + a_1' b_2 = -M_3$.

NETWORK 18



(To simulate a short symmetrical line or circuit)

Symmetrical Section of Line or Circuit:

$$X = \text{open-circuit impedance};$$

$$Y = \text{short-circuit impedance};$$

$$\tanh^{-1}\sqrt{Y/X} = \text{propagation length};$$

$$\sqrt{XY} = \text{iterative impedance}.$$

Simulating Network:

$$z_a = \sqrt{XY}\tanh^{-1}\sqrt{Y/X};$$

$$z_b = \sqrt{XY}/\tanh^{-1}\sqrt{Y/X};$$

$m_1 = .45737;$ $m_2 = .14456;$ $m_1' = .04263;$ $m_2' = .92403.$

The impedances z_a and z_b are to be realized in desired frequency ranges, more or less approximately, by comparatively simple physical networks.

Transmission of Information[1]

By R. V. L. HARTLEY

SYNOPSIS: A quantitative measure of "information" is developed which is based on physical as contrasted with psychological considerations. How the rate of transmission of this information over a system is limited by the distortion resulting from storage of energy is discussed from the transient viewpoint. The relation between the transient and steady state viewpoints is reviewed. It is shown that when the storage of energy is used to restrict the steady state transmission to a limited range of frequencies the amount of information that can be transmitted is proportional to the product of the width of the frequency-range by the time it is available. Several illustrations of the application of this principle to practical systems are included. In the case of picture transmission and television the spacial variation of intensity is analyzed by a steady state method analogous to that commonly used for variations with time.

WHILE the frequency relations involved in electrical communication are interesting in themselves, I should hardly be justified in discussing them on this occasion unless we could deduce from them something of fairly general practical application to the engineering of communication systems. What I hope to accomplish in this direction is to set up a quantitative measure whereby the capacities of various systems to transmit information may be compared. In doing this I shall discuss its application to systems of telegraphy, telephony, picture transmission and television over both wire and radio paths. It will, of course, be found that in very many cases it is not economically practical to make use of the full physical possibilities of a system. Such a criterion is, however, often useful for estimating the possible increase in performance which may be expected to result from improvements in apparatus or circuits, and also for detecting fallacies in the theory of operation of a proposed system.

Inasmuch as the results to be obtained are to represent the limits of what may be expected under rather idealized conditions, it will be permissible to simplify the discussion by neglecting certain factors which, while often important in practice, have the effect only of causing the performance to fall somewhat further short of the ideal. For example, external interference, which can never be entirely eliminated in practice, always reduces the effectiveness of the system. We may, however, arbitrarily assume it to be absent, and consider the limitations which still remain due to the transmission system itself.

In order to lay the groundwork for the more practical applications of these frequency relationships, it will first be necessary to discuss a few somewhat abstract considerations.

[1] Presented at the International Congress of Telegraphy and Telephony, Lake Como, Italy, September 1927.

The Measurement of Information

When we speak of the capacity of a system to transmit information we imply some sort of quantitative measure of information. As commonly used, information is a very elastic term, and it will first be necessary to set up for it a more specific meaning as applied to the present discussion. As a starting place for this let us consider what factors are involved in communication; whether conducted by wire, direct speech, writing, or any other method. In the first place, there must be a group of physical symbols, such as words, dots and dashes or the like, which by general agreement convey certain meanings to the parties communicating. In any given communication the sender mentally selects a particular symbol and by some bodily motion, as of his vocal mechanism, causes the attention of the receiver to be directed to that particular symbol. By successive selections a sequence of symbols is brought to the listener's attention. At each selection there are eliminated all of the other symbols which might have been chosen. As the selections proceed more and more possible symbol sequences are eliminated, and we say that the information becomes more precise. For example, in the sentence, "Apples are red," the first word eliminates other kinds of fruit and all other objects in general. The second directs attention to some property or condition of apples, and the third eliminates other possible colors. It does not, however, eliminate possibilities regarding the size of apples, and this further information may be conveyed by subsequent selections.

Inasmuch as the precision of the information depends upon what other symbol sequences might have been chosen it would seem reasonable to hope to find in the number of these sequences the desired quantitative measure of information. The number of symbols available at any one selection obviously varies widely with the type of symbols used, with the particular communicators and with the degree of previous understanding existing between them. For two persons who speak different languages the number of symbols available is negligible as compared with that for persons who speak the same language. It is desirable therefore to eliminate the psychological factors involved and to establish a measure of information in terms of purely physical quantities.

Elimination of Psychological Factors

To illustrate how this may be done consider a hand-operated submarine telegraph cable system in which an oscillographic recorder traces the received message on a photosensitive tape. Suppose the

sending operator has at his disposal three positions of a sending key which correspond to applied voltages of the two polarities and to no applied voltage. In making a selection he decides to direct attention to one of the three voltage conditions or symbols by throwing the key to the position corresponding to that symbol. The disturbance transmitted over the cable is then the result of a series of conscious selections. However, a similar sequence of arbitrarily chosen symbols might have been sent by an automatic mechanism which controlled the position of the key in accordance with the results of a series of chance operations such as a ball rolling into one of three pockets.

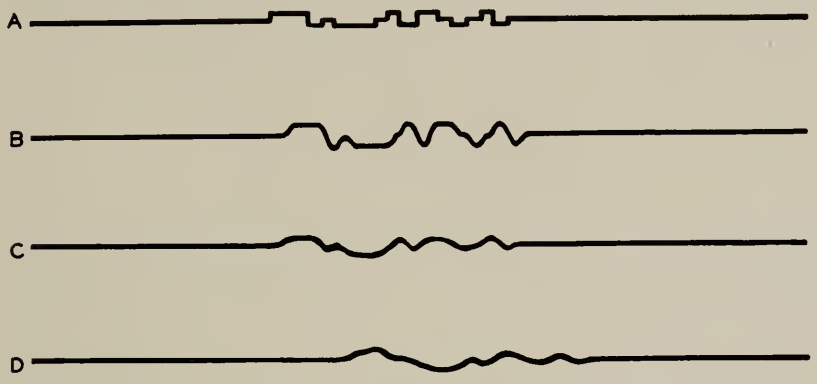


Fig. 1

Owing to the distortion of the cable the results of the various selections as exhibited to the receiver by the recorder trace are not as clearly distinguishable as they were in the positions of the sending key. Fig. 1 shows at *A* the sequence of key positions, and at *B*, *C* and *D* the traces made by the recorder when receiving over an artificial cable of progressively increasing length. For the shortest cable *B* the reconstruction of the original sequence is a simple matter. For the intermediate length *C*, however, more care is needed to distinguish just which key position a particular part of the record represents. In *D* the symbols have become hopelessly indistinguishable. The capacity of a system to transmit a particular sequence of symbols depends upon the possibility of distinguishing at the receiving end between the results of the various selections made at the sending end. The operation of recognizing from the received record the sequence of symbols selected at the sending end may be carried out by those of us who are not familiar with the Morse code. We would do this equally well for a sequence representing a consciously chosen message and for one sent out by the automatic selecting device already referred

to. A trained operator, however, would say that the sequence sent out by the automatic device was not intelligible. The reason for this is that only a limited number of the possible sequences have been assigned meanings common to him and the sending operator. Thus the number of symbols available to the sending operator at certain of his selections is here limited by psychological rather than physical considerations. Other operators using other codes might make other selections. Hence in estimating the capacity of the physical system to transmit information we should ignore the question of interpretation, make each selection perfectly arbitrary, and base our result on the possibility of the receiver's distinguishing the result of selecting any one symbol from that of selecting any other. By this means the psychological factors and their variations are eliminated and it becomes possible to set up a definite quantitative measure of information based on physical considerations alone.

Quantitative Expression for Information

At each selection there are available three possible symbols. Two successive selections make possible 3^2, or 9, different permutations or symbol sequences. Similarly n selections make possible 3^n different sequences. Suppose that instead of this system, in which three current values are used, one is provided in which any arbitrary number s of different current values can be applied to the line and distinguished from each other at the receiving end. Then the number of symbols available at each selection is s and the number of distinguishable sequences is s^n.

Consider the case of a printing telegraph system of the Baudot type, in which the operator selects letters or other characters each of which when transmitted consists of a sequence of symbols (usually five in number). We may think of the various current values as primary symbols and the various sequences of these which represent characters as secondary symbols. The selection may then be made at the sending end among either primary or secondary symbols. Let the operator select a sequence of n_2 characters each made up of a sequence of n_1 primary selections. At each selection he will have available as many different secondary symbols as there are different sequences that can result from making n_1 selections from among the s primary symbols. If we call this number of secondary symbols s_2, then

$$s_2 = s^{n_1}. \tag{1}$$

For the Baudot System

$$s_2 = 2^5 = 32 \text{ characters}. \tag{2}$$

The number of possible sequences of secondary symbols that can result from n_2 secondary selections is

$$s_2^{n_2} = s^{n_1 n_2}. \tag{3}$$

Now $n_1 n_2$ is the number n of selections of primary symbols that would have been necessary to produce the same sequence had there been no mechanism for grouping the primary symbols into secondary symbols. Thus we see that the total number of possible sequences is s^n regardless of whether or not the primary symbols are grouped for purposes of interpretation.

This number s^n is then the number of possible sequences which we set out to find in the hope that it could be used as a measure of the information involved. Let us see how well it meets the requirements of such a measure.

For a particular system and mode of operation s may be assumed to be fixed and the number of selections n increases as the communication proceeds. Hence with this measure the amount of information transmitted would increase exponentially with the number of selections and the contribution of a single selection to the total information transmitted would progressively increase. Doubtless some such increase does often occur in communication as viewed from the psychological standpoint. For example, the single word "yes" or "no," when coming at the end of a protracted discussion, may have an extraordinarily great significance. However, such cases are the exception rather than the rule. The constant changing of the subject of discussion, and even of the individuals involved, has the effect in practice of confining the cumulative action of this exponential relation to comparatively short periods.

Moreover we are setting up a measure which is to be independent of psychological factors. When we consider a physical transmission system we find no such exponential increase in the facilities necessary for transmitting the results of successive selections. The various primary symbols involved are just as distinguishable at the receiving end for one primary selection as for another. A telegraph system finds one ten-word message no more difficult to transmit than the one which preceded it. A telephone system which transmits speech successfully now will continue to do so as long as the system remains unchanged. In order then for a measure of information to be of practical engineering value it should be of such a nature that the information is proportional to the number of selections. The number of possible sequences is therefore not suitable for use directly as a measure of information.

We may, however, use it as the basis for a derived measure which does meet the practical requirements. To do this we arbitrarily put the amount of information proportional to the number of selections and so choose the factor of proportionality as to make equal amounts of information correspond to equal numbers of possible sequences. For a particular system let the amount of information associated with n selections be

$$H = Kn, \tag{4}$$

where K is a constant which depends on the number s of symbols available at each selection. Take any two systems for which s has the values s_1 and s_2 and let the corresponding constants be K_1 and K_2. We then define these constants by the condition that whenever the numbers of selections n_1 and n_2 for the two systems are such that the number of possible sequences is the same for both systems, then the amount of information is also the same for both; that is to say, when

$$s_1^{n_1} = s_2^{n_2}, \tag{5}$$

$$H = K_1 n_1 = K_2 n_2, \tag{6}$$

from which

$$\frac{K_1}{\log s_1} = \frac{K_2}{\log s_2}. \tag{7}$$

This relation will hold for all values of s only if K is connected with s by the relation

$$K = K_0 \log s, \tag{8}$$

where K_0 is the same for all systems. Since K_0 is arbitrary, we may omit it if we make the logarithmic base arbitrary. The particular base selected fixes the size of the unit of information. Putting this value of K in (4),

$$H = n \log s \tag{9}$$

$$= \log s^n. \tag{10}$$

What we have done then is to take as our practical measure of information the logarithm of the number of possible symbol sequences.

The situation is similar to that involved in measuring the transmission loss due to the insertion of a piece of apparatus in a telephone system. The effect of the insertion is to alter in a certain ratio the power delivered to the receiver. This ratio might be taken as a measure of the loss. It is found more convenient, however, to take the logarithm of the power ratio as a measure of the transmission loss.

If we put n equal to unity, we see that the information associated with a single selection is the logarithm of the number of symbols available; for example, in the Baudot System referred to above, the number s of primary symbols or current values is 2 and the information content of one selection is log 2; that of a character which involves 5 selections is 5 log 2. The same result is obtained if we regard a character as a secondary symbol and take the logarithm of the number of these symbols, that is, log 2^5, or 5 log 2. The information associated with 100 characters will be 500 log 2. The numerical value of the information will depend upon the system of logarithms used. Increasing the number of current values from 2 to say 10, that is, in the ratio 5, would increase the information content of a given number of selections in the ratio $\frac{\log 10}{\log 2}$, or 3.3. Its effect on the rate of transmission will depend upon how the rate of making selections is affected. This will be discussed later.

When, as in the case just considered, the secondary symbols all involve the same number of primary selections, the relations are quite simple. When a telegraph system is used which employs a non-uniform code they are rather more complicated. A difficulty, more apparent than real, arises from the fact that a given number of secondary or character selections may necessitate widely different numbers of primary selections, depending on the particular characters chosen. This would seem to indicate that the values of information deduced from the primary and secondary symbols would be different. It may easily be shown, however, that this does not necessarily follow.

If the sender is at all times free to choose any secondary symbol, he may make all of his selections from among those containing the greatest number of primary symbols. The secondary symbols will then all be of equal length, and, just as for the uniform code, the number of primary symbols will be the product of the number of characters by the maximum number of primary selections per character. If the number of primary selections for a given number of characters is to be kept to some smaller value than this, some restriction must be placed on the freedom of selection of the secondary symbols. Such a restriction is imposed when, in computing the average number of dots per character for a non-uniform code, we take account of the average frequency of occurrence of the various characters in telegraph messages. If this allotted number of dots per character is not to be exceeded in sending a message, the operator must, on the average, refrain from selecting the longer characters more often than their average rate of occurrence. In the language

of the present discussion we would say that for certain of the n_2 secondary selections the value of s_2, the number of secondary symbols, is so reduced that a summation of the information content over all the characters gives a value equal to that derived from the total number of primary selections involved. This may be written

$$\sum_{1}^{n_2} \log s_2 = n \log s, \tag{11}$$

where n is the total number of primary symbols or dot lengths assigned to n_2 characters. This suggests that the primary symbols furnish the most convenient basis for evaluating information.

The discussion so far has dealt largely with telegraphy. When we attempt to extend this idea to other forms of communication certain generalizations need to be made. In speech, for example, we might assume the primary selections to represent the choice of successive words. On that basis s would represent the number of available words. For the first word of a conversation this would correspond to the number of words in the language. For subsequent selections the number would ordinarily be reduced because subsequent words would have to combine in intelligible fashion with those preceding. Such limitations, however, are limitations of interpretation only and the system would be just as capable of transmitting a communication in which all possible permutations of the words of the language were intelligible. Moreover, a telephone system may be just as capable of transmitting speech in one language as in another. Each word may be spoken in a variety of ways and sung in a still greater variety. This very large amount of information associated with the selection of a single spoken word suggests that the word may better be regarded as a secondary symbol, or sequence of primary symbols. Let us see where this point of view leads us.

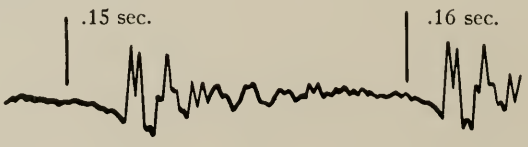


Fig. 2

The actual physical embodiment of the word consists of an acoustic or electrical disturbance which may be expressed as a magnitude-time function as in Fig. 2, which shows an oscillographic record of a speech sound. Such functions are also typical of other modes of communication, as will be discussed in more detail later. We have then to examine the ability of such a continuous function to convey informa-

tion. Obviously over any given time interval the magnitude may vary in accordance with an infinite number of such functions. This would mean an infinite number of possible secondary symbols, and hence an infinite amount of information. In practice, however, the information contained is finite for the reason that the sender is unable to control the form of the function with complete accuracy, and any distortion of its form tends to cause it to be confused with some other function.

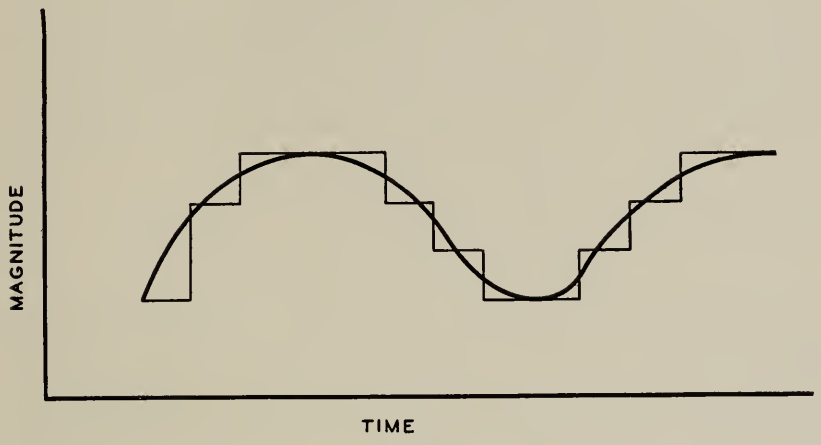


Fig. 3

A continuous curve may be thought of as the limit approached by a curve made up of successive steps, as shown in Fig. 3, when the interval between the steps is made infinitesimal. An imperfectly defined curve may then be thought of as one in which the interval between the steps is finite. The steps then represent primary selections. The number of selections in a finite time is finite. Also the change made at each step is to be thought of as limited to one of a finite number of values. This means that the number of available symbols is kept finite. If this were not the case, the curve would be defined with complete exactness at each of the steps, which would mean that an observation made at any one step would offer the possibility of distinguishing among an infinite number of possible values. The following illustration may serve to bring out the relation between the discrete selections and the corresponding continuous curve. We may think of a bicycle equipped with a peculiar type of steering device which permits the rider to set the front wheel in only a limited number of fixed positions. On such a machine he attempts to ride in such a manner that the front wheel shall follow an irregularly

curved line. The accuracy with which he is able to accomplish this will depend upon how far he goes between adjustments of the steering mechanism and upon the number of positions in which he is able to set it.

By this more or less artificial device the continuous magnitude-time function as used in telephony is made subject to the same type of treatment as the succession of discrete selections involved in telegraphy.

Rate of Communication

So far then we have derived an expression for the information content of the symbols at the sending end and have shown that we may evaluate a transmission system in terms of how well the wave as received over it permits distinguishing between the various possible symbols which are available for each selection. Let us consider next how the distortion of the system limits the rate of selection for which these distinctions between symbols may be made with certainty.

Limitation by Intersymbol Interference

We shall assume the system to be free from external interference and to be such that its current-voltage relations are linear. In such a system the form of the transmitted wave may be altered due to the storage of energy in reactive elements such as inductances and capacities, and its subsequent release. To evaluate the effect of such distortion in making it impossible to determine correctly which one of the available symbols had been selected, we may think of this distortion in terms of "intersymbol interference." In order to determine the result of any one selection an observation is made at such time that the disturbance resulting from that selection has its maximum effect at the receiving end. Superposed on this effect there will be a disturbance which is the resultant of the effects of all the other symbols as prolonged by the storage of energy in the system. This resultant superposed disturbance is what is meant by intersymbol interference. Obviously if this disturbance is greater than half the difference between the effects produced by two of the values available for selection at the sending end, the wave resulting from one of those values will be taken as representing the other. Thus a criterion for successful transmission is that in no case shall the intersymbol interference exceed half the difference between the values of the wave at the receiving end which correspond to the selection of different values at the sending end.

Obviously the magnitude of the intersymbol interference which affects any one symbol depends on the particular sequence of symbols

which precedes it. However, it is always possible for the sending operator so to make his selections that any one selection is preceded by that sequence which causes the maximum possible interference. Hence every selection must be separated from those preceding it by at least a certain interval which is determined by the worst condition of interference. If longer intervals than this are used, the transmission is unnecessarily retarded. Hence to secure the maximum rate of transmission the selection should be made at a constant rate. It might appear at first sight that the selections could be made at shorter intervals near the beginning of the message where there are fewer preceding symbols to cause interference. This assumes, however, that the system has previously been idle. Actually the previous user may have finished his message with that sequence which causes maximum intersymbol interference.

Relation to Damping Constant

How the intersymbol interference limits the rate of communication over the system depends upon the properties of the particular system. The relations involved are very complex, and no attempt will be made to obtain a complete or rigorous solution of the problem. We may, however, by treating a very simple case, arrive at an interesting relation. Consider a resistance in series with a capacity. Let one terminal be connected to one terminal of a battery made up of a very large number of cells of negligible internal resistance. Let the other terminal be connected to the battery through a switch. This switch is so arranged that by pressing any one of s keys the circuit terminal may be moved up along the battery by any number of cells from zero to $s - 1$. Let the sending operator make selections among the s keys at regular intervals, and let the receiving operator observe the current through the resistance. The most advantageous time for this observation is at the instant [1] at which the key is pressed, since the current has then its maximum value. The finest distinction to be made by the receiving operator is that between two currents which result from battery changes that differ from each other by one cell. The difference between two such currents is equal to the initial current which flows when one cell is introduced into the circuit. This is

$$i_s = \frac{E}{R}, \tag{12}$$

where E is the electromotive force of one cell and R the resistance of the circuit.

[1] Results identical with those which follow may be obtained if he observes the average current over a period beginning when the key is pressed and lasting not longer than the interval between selections.

The intersymbol interference will consist of currents resulting from all of the preceding symbols. The contribution of any one symbol will depend on its size, that is, on the number of added cells it represents, and on how long it preceded the symbol in question. For a given rate of selection the resultant of these contributions will be a maximum for that particular sequence of symbols for which at every selection preceding the one in question the operator had selected the largest

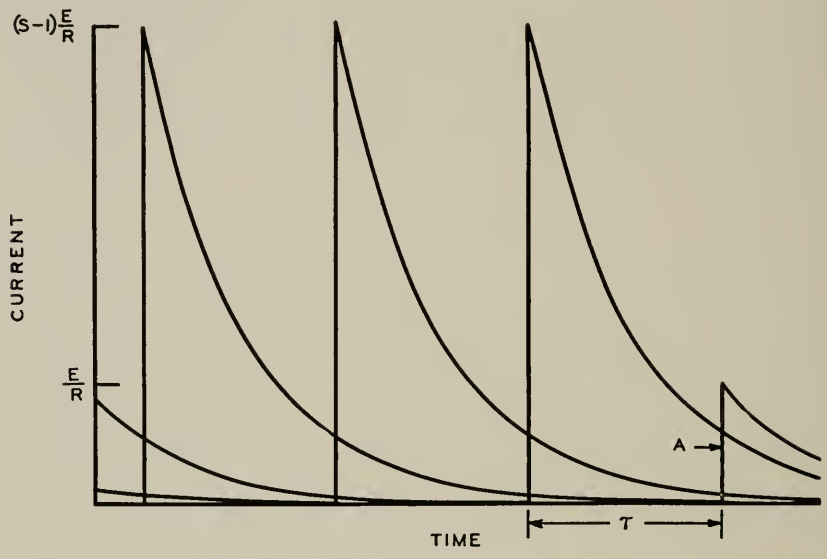


Fig. 4

possible symbol, that is, a voltage change of $(s - 1)E$. The form of the received current is then as shown on Fig. 4, where A represents the disturbed symbol. The curves are drawn for s equal to five. The current resulting from one such change occurring at time zero is

$$i = (s - 1)\frac{E}{R}e^{-\alpha t}, \tag{13}$$

where the damping constant,

$$\alpha = \frac{1}{RC}\cdot \tag{14}$$

If the interval between selections is τ, then the time during which any one interfering current is damped out before it makes its contribution to the interference with the disturbed symbol is $q\tau$ where q is the number of selections by which it precedes the disturbed symbol. The magnitude of its contribution is therefore from (13)

$$i_q = (s - 1)\frac{E}{R}e^{-q\alpha\tau}. \tag{15}$$

If we sum this expression for all values of q from one to infinity, we get the combined effect of all the preceding symbols, that is, the intersymbol interference. Calling this i_i,

$$i_i = (s - 1)\frac{E}{R}\sum_{q=1}^{q=\infty} e^{-q\alpha\tau} \tag{16}$$

$$= (s - 1)\frac{E}{R}\frac{1}{e^{\alpha\tau} - 1}. \tag{17}$$

This obviously increases as the interval τ between selections is decreased. If this interval is made small enough, the intersymbol interference may cause confusion between symbols. Since the interference is here always of one sign it can cause confusion only when it becomes as large as the minimum difference, i_s, between symbols. Placing these two quantities equal, we get from (12) and (17) as the minimum permissible value of τ,

$$\tau_1 = \frac{\log s}{\alpha}. \tag{18}$$

The maximum number, n, of selections that may be made in t seconds is given by

$$n = \frac{t}{\tau_1}. \tag{19}$$

From (18) and (19)

$$\frac{n \log s}{t} = \alpha. \tag{20}$$

Here the numerator is, in accordance with our measure of information, the amount of information contained in n selections, so the left-hand member is the information per unit time or the rate of communication. This is equal to the damping constant of the circuit. We therefore conclude that for this particular case the possible rate of communication is fixed solely by the damping constant of the circuit and is independent of the number of symbols available at each selection. It is, of course, true that the larger this number the more susceptible will the system be to the effects of external interference.

Probably the practical system which most nearly approaches this idealized one is the non-loaded submarine telegraph cable when operated at such low speeds that its inductance may be neglected. It is of considerable historical interest to note that Lord Kelvin's

study of such cables led him to the conclusion that the extent to which the cable limited the dotting speed was given by KR, that is, the product of the total capacity and total resistance. Had he stated his results in terms of permissible speed he would have had the reciprocal of this quantity, which corresponds very closely to the damping constant which we arrived at as a measure of the rate of communication. It should be noted, however, that his consideration was limited to a fixed number of symbols, and did not involve the relation here developed between this number and the dotting speed.

The more complicated systems are similar to the simple case just treated in that the contribution of any one symbol, a, to the interference with any other symbol, b, is determined by the free vibration of the system which results from the change applied to it in the production of symbol a. This free vibration, instead of being expressible by a single exponential function as in the case just considered, may be the resultant of a large number of more or less damped oscillatory components corresponding to the various natural modes of the system. The total interference with any one symbol is the resultant of a series of these complex vibrations, one for each interfering symbol. The instantaneous values of the various components of the interference are so dependent upon their phases at the particular instant of observation that it is difficult to draw any general conclusions as to the magnitude of the total interference. It is equally difficult therefore to draw any general conclusions as to the relation between the rate of transmission over a particular circuit and the number of available symbols.

Relation to Storage of Energy

Even though for any one system there exists a number of available symbols for which the rate of communication is greater than for any other number, it is still possible to make a generalization with respect to the storage of energy in the system and its effect on the rate of transmission which is of considerable practical importance.

Each of the natural modes of vibration of a linear system has the general form

$$i = Ae^{-\alpha t} \cos (\omega t - \theta), \tag{21}$$

where the natural frequency ω and damping constant α are characteristic of the system and the amplitude A and phase θ depend on the conditions of excitation. Wherever the time appears in this expression it is multiplied by either the damping constant α or the frequency ω. Consequently if both α and ω be changed, say in the ratio k, the instantaneous value of this mode of vibration in the new system will

be the same at a time t/k that it was at time t in the original system. If the same change is made in the damping constant and frequency of each one of the modes of free vibration, their resultant, or the wave set up by any one symbol, will also be so changed that any particular value occurs at time t/k instead of t. Suppose also that the interval τ between selections be changed to τ/k. Then any two symbols originally separated by a time t_1 will be separated by t_1/k. The value of the interfering wave at the time t_1/k when the disturbed symbol occurs will be the same as it was at the corresponding time t_1 when it occurred in the original system. Hence the contribution of this wave to the intersymbol interference is unchanged. Since this relation holds for all of the interfering symbols, the total intersymbol interference remains unchanged, and so the number of possible symbols that may be distinguished is unaltered. The rate of making selections is changed in the ratio k, and hence the maximum rate of communication is changed in the same ratio as the damping constants and natural frequencies.

Now let us consider what physical changes must be made in the system to bring about the assumed changes in the damping constants and natural frequencies of the various modes. Take the simple case of an inductance, capacity and resistance connected in series. Here we have the well-known relations

$$\alpha = \frac{R}{2L}, \tag{22}$$

$$\omega = \sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}. \tag{23}$$

If R remains fixed and L is changed to L/k, α becomes $k\alpha$. If, in addition, C is changed to C/k, ω becomes $k\omega$. What we have done is to leave the energy-dissipating element, R, unchanged and change both the energy-storing elements, L and C, in the inverse ratio in which the rate of communication is changed. For more complicated systems the expressions for α and ω are correspondingly complicated. In every case, however, it will be found that if all of the dissipating and storing elements are treated as in the simple case just considered all of the damping constants and natural frequencies will be similarly altered. Where mechanical systems are concerned we are to substitute for electrical resistances their mechanical equivalents, and for inductances and capacities, inertias and compliances. This generalization that a proportionate change in all of the energy-storing elements of the system with no accompanying change in the dissipating elements

produces an inverse change in the possible rate of transmission will be used later.

Steady State and Transient Viewpoints

So far very little has been said about frequencies, and in fact nothing in the sense of the term in which our results are to be stated. By this I mean the use of the word "frequency" as applied to an alternating current or other sinusoidal disturbance in the so-called "steady state." The steady state viewpoint has proven very useful in certain branches of communication, notably telephony. During the past few years much progress has been made in establishing relations between steady state phenomena and what might be called transient phenomena of the sort which we have just been discussing. Before proceeding with the main argument I shall attempt to review in non-mathematical language the relationship of these two points of view to each other.

As its name implies, steady state analysis deals with continuing conditions. If a sustained sinusoidal electromotive force be applied at the sending end of a system, a sinusoidal current of the same frequency flows at the receiving end. The vector ratio of the received current to the sending electromotive force is known as the transfer admittance of the system at that frequency. It is assumed ideally that the driving electromotive force has been acting from the beginning of time, and practically that it has been acting so long that the results are indistinguishable from what would be obtained in the ideal case. The absolute magnitude of the transfer admittance gives the amplitude of the received current which results from a driving electromotive force of unit amplitude, and its phase angle gives the phase of the current relative to that of the driving electromotive force. The curves which represent this amplitude and phase as functions of the frequency constitute a steady state description of the transmission properties of the system. For a system which is free from energy storage such as a circuit containing resistances only, the transfer admittance is the same for all frequencies. The amplitude-frequency curve is a horizontal line whose position depends on the magnitude and arrangement of the resistances while the phase-frequency curve coincides with the frequency axis. The storage of energy in the system and its subsequent release cause the admittance-frequency curves to take other forms. If the only storage is that which occurs in a dissipationless medium, a condition which is approximated when a sound wave traverses the open air, the only effect is to make the phase-frequency curve a straight line passing through the origin and having a slope proportional to the time of transmission through the

medium. Other forms of energy storage give to the admittance-frequency curves shapes which are characteristic of the particular system. This alteration of these curves is commonly spoken of as frequency distortion. Their form may in most cases be deduced fairly readily from the values of the energy-storing and energy-dissipating elements of the system. This fact makes such a description of the system particularly useful for design purposes.

This physical description is, of course, useful as a criterion of performance only in so far as it can be related to the satisfactoriness with which the system performs its primary function of transmitting information. In the case of telephony it has been found practical to establish such a correlation by purely empirical means. Until fairly recently adequate results have been obtained by considering the amplitude-frequency function only. With the use of lines of increasing length and with increasingly severe standards of performance it is coming to be necessary to take account of the phase-frequency function as well.

In attempting to extend this method of treatment to telegraphy it was not found desirable to establish the correlation between steady state properties and overall performance by purely empirical methods. One reason for this was that a considerable fund of information has been accumulated with reference to the correlation between the overall performance and the transient properties of the system. A correlation therefore between steady state and transient properties would offer a means of bringing this empirical information to bear on the design of apparatus and systems on a steady state basis. For bridging this gap between steady state and transient phenomena there was already available one arch in the form of the Fourier Integral. This integral may be thought of as a mathematical fiction for expressing a transient phenomenon in terms of steady state phenomena. It permits the determination, for any magnitude-time function, of the relative amplitudes and the phases of an infinite succession of sustained sinusoids whose resultant is at any instant equal to the magnitude of the function at that instant. The amplitudes of the sinusoids are infinitesimal and the frequencies of successive components differ from each other by infinitesimal increments. The relative amplitudes and the phases of these components expressed as functions of the frequency constitute a steady state description of the magnitude-time function.

Suppose then we have given the magnitude-time function representing an impressed transient driving force and wish to obtain the magnitude-time function of the received current. We deduce the steady state description of the driving force, modify the amplitudes

and phases of its various components in accordance with the known admittance-frequency functions of the system, and obtain the amplitude- and phase-frequency curves which represent the steady state description of the received current. From these we deduce the magnitude-time function representing the received wave.

A slightly different point of view, however, leads to results which fit in better with our method of measuring information. We may apply the method just outlined to deduce the magnitude-time function which results when the applied wave consists merely of an instantaneous change in a steadily applied electromotive force from one value which may be zero to another value differing from it by one unit. The resulting wave form is characteristic of the system and has been called by J. R. Carson its indicial admittance. We may think of this as a transient description of the system. If we regard a continuously varying applied wave as being formed by a succession of steps, we may think of the received wave at any instant as being the resultant of a series of waves each corresponding to a single step. The wave form of each is that of the indicial admittance, its magnitude is proportional to the size of the particular step and its location on the time axis is determined by the time at which the particular step or selection was made. When the steps are made infinitely close together, a summation of these components becomes a process of integration whereby the resulting magnitude-time function may be accurately determined from the applied function. For the incompletely determined waves involved in communication where the separation of the steps is finite a corresponding summation of the indicial admittance curves resulting from all selections other than the one being observed gives a measure of the intersymbol interference.

Still another viewpoint, while it has, perhaps, less direct application to the present problem, is of interest in that it brings out the significance from the transient standpoint of the steady state characteristics of the system. If we take as the applied wave a mathematical impulse, that is to say, a disturbance which lasts for an infinitesimal time, we find that the amplitudes of its steady state components are the same for all frequencies. If the impulse occurs at zero time the phase-frequency curve coincides with the axis of frequency, and if not it is a straight line through the origin whose slope is proportional to the time of occurrence. In order to find the current resulting from such an impulse applied at zero time we multiply the constant amplitude-frequency curve of its steady state components by the amplitude-frequency curve of the system and obtain as the amplitude-frequency curve of the received wave a function of the same form as the ampli-

tude-frequency curve of the system. Similarly we add to the phase-frequency curve of the impressed wave, which is zero at all frequencies, the phase-frequency curve of the system and obtain for the received wave a phase-frequency curve identical with that of the system. The corresponding magnitude-time function gives the instantaneous value of the received current resulting from the impressed impulse. Thus we see that the steady state transfer admittance of a system is identical with the steady state description of the wave which is received over the system when it is subjected to an impulsive driving force. Once the form of this received wave is known the received wave resulting from any applied wave may be deduced by assuming the applied wave to consist of an infinite succession of impulses infinitesimally close together whose magnitudes vary with time in accordance with the given magnitude-time function. Methods for integrating the effect of this infinite succession of responses to impulses so as to obtain the transmitted wave have been developed.

From this review it is evident that the so-called frequency distortion and transient distortion are merely two methods of describing the same changes in wave form which result from the storage of energy in parts of the transmission system.

Significance of Product of Frequency-Range by Time

Distortion of this sort with its accompanying intersymbol interference may be unavoidable in the design of the system, or it may be deliberately introduced. The use of electrical filters to obtain multi-

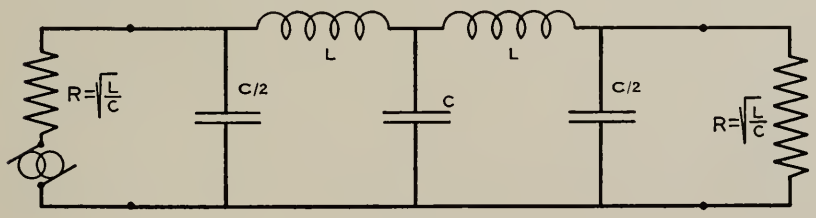


Fig. 5

plex operation, as in carrier systems, is an example of its deliberate use. Consider the effect of introducing a low pass filter, as shown in Fig. 5, into an otherwise distortionless transmission system. If the impedances of the circuits to which the filter is connected are approximately pure resistances of the values indicated in the figure, steady state frequencies above a critical value known as the cut-off frequency are so reduced as to be made practically negligible, while frequencies below this value are transmitted with very little distortion. The

transient distortion corresponding to this steady state distortion must result in intersymbol interference; hence it places a limit on the rate at which distinguishable symbols may be selected, that is, on the rate of transmitting information.

It does not necessarily follow, however, that the rate of transmission with such a system is the maximum attainable for systems whose transmission is limited to the frequency-range determined by the cut-off of the filter. It is conceivable that by the introduction of additional energy-storing elements the transfer admittance curves for frequencies within the transmitted range may be altered in such a way as to reduce the total intersymbol interference and so permit an increased rate of selection. The maximum rate of transmission of information which can be secured by such methods represents the maximum rate corresponding to that range of frequencies.

Let us consider next the way in which this possible rate of transmission varies with the cut-off frequency of the filter. The theory of filter design teaches us that the cut-off frequency may be changed without altering the required terminating resistances if we change all inductances and capacities in the inverse ratio of the desired change in cut-off frequency. Suppose this change in energy-storing elements, with no change in dissipative elements, is made not only for the filter but for the entire system. We have already seen that such a modification changes the rate of transmission in the inverse ratio of the change in energy-storing elements; that is, in the direct ratio of the change in cut-off frequency in the present case. That the new rate is the maximum for the new frequency-range is evident when we consider that the transfer admittance curves of the new system bear the same relation to its cut-off frequency as held in the original system.

This brings us to the important conclusion that the maximum rate at which information may be transmitted over a system whose transmission is limited to frequencies lying in a restricted range is proportional to the extent of this frequency-range. From this it follows that *the total amount of information which may be transmitted over such a system is proportional to the product of the frequency-range which it transmits by the time during which it is available for the transmission.* This product of transmitted frequency-range by time available is the quantitative criterion for comparing transmission systems to which I referred at the beginning of this discussion. The significance of this criterion can perhaps best be brought out by applying it to some typical situations.

Fitting the Messages to the Lines

To facilitate this discussion it seems desirable to introduce and explain a few terms. For transmitting a sequence of symbols various sorts of media may be available, such as a wire line, an air path, as in direct speech, or the ether, as in radio communication. For convenience we shall group all of these under the general name of "line." Each such medium is generally characterized by a range of frequencies over which transmission may be carried on with reasonable freedom from distortion and external interference. This will be called the "line-frequency-range." Similarly the symbol sequences corresponding to the various modes of communication such as telegraph and telephone, will be designated as "messages." Each of these will, in general, be characterized by a "message frequency-range." This may be thought of as being determined by the frequency-range of that line which will just transmit the type of message satisfactorily, or we may think of it as that part of the frequency scale within which it is necessary to preserve the steady state components of the message wave in order to permit distinguishing the various symbols as they appear in the transmitted wave.

When we set up practical communication systems it is often found that the message-frequency-range and the line-frequency-range do not coincide either in magnitude or in position on the frequency scale. If then we are to make use of the full transmission capacity of the line, or lines, we must introduce means for altering the frequency-ranges required by the messages. Two such means are available, which together offer the theoretical possibility of accomplishing the desired end of making the message-frequency-ranges fit the available line-frequency-ranges.

The process of modulation so widely used in radio systems and in carrier transmission over wires makes it possible to shift the frequency-range of any message to a new location on the frequency scale without altering the width of the range. This follows at once from the well-known fact that the steady state description of the wave which results from the modulation of a carrier wave by a symbol wave includes a pair of side-bands in each of which there is a component corresponding to each steady state component of the original wave. The frequency of each component of the side-band differs from the carrier frequency by the frequency of the corresponding component of the symbol wave. The elimination of one of these side-bands results in a wave which retains the information embodied in the original symbol wave and occupies a frequency-range of the same width

as the original but displaced to a new position on the frequency scale determined by the carrier frequency. The interval which must be allowed between these displaced messages in carrier operation is determined by the selectivity of the filters which are available for their separation. The imperfection of practical filters tends to make the message-frequency-range which may be transmitted less than the line-frequency-range which the messages occupy. The time for which the line is used to transmit a given amount of information is the same as the duration of the message conveying it. Thus the sum of the products of frequency-range by time for the messages is always equal to or less than the corresponding sum of the products of line-frequency-range by time.

In case the line-range available is less than the message-range, as would be the case in attempting to transmit speech over a submarine telegraph cable, it is still possible, if enough lines are available, to accomplish the transmission. The message wave may, by suitable filters, be separated into a plurality of waves each made up of those components of the original which lie in a portion of the message-range which is no wider than the line-range. Each of these portions of the message may then, by modulation, be transferred down to the frequency-range of the line and each transmitted over a separate line. A reversal of the process at the receiving end restores the original message.

While it is theoretically possible, if enough messages and lines are available, to fit the message-ranges to the line-ranges by modulation and subdivision of message-frequency-ranges, it is not always practical. It is sometimes more desirable to utilize the second method of transformation already referred to. This consists in making a record of the symbol sequence and reproducing it at a different speed in order to secure the wave used in transmission. The tape used in sending telegraph messages may be used in this manner. Here the symbol sequence represents a series of selections of secondary symbols. These selections are made at a rate at which it is convenient for the operator to manipulate the keys of the tape-punching machine. The electric wave impressed on the line by the holes in the tape represents a corresponding sequence of primary symbols. The rate at which these are applied to the line is determined by the velocity of the tape in reproduction. Since for a given number of different primary symbols the frequency-range required is proportional to the rate of making selections, it is obvious that the frequency-range of the message as reproduced from the tape may be made to fit whatever line-frequency-range is available, at least so far as width of the range is concerned.

Modulation may, of course, be necessary to bring the message-range to the proper part of the frequency scale. The time required for the reproduction of a message involving a given number of selections varies inversely as the velocity of the tape in reproduction, and therefore also inversely as the frequency-range required by the reproduced sequence. Thus the product of frequency-range by time for the reproduced message, which is also the required product for the line, is independent of the rate of reproduction, and depends only on the information content of the message in its original form.

In case the available line range calls for reproduction at a considerably increased speed a single operator cannot conveniently keep the sending apparatus supplied with tape. Multiplex operation may then be employed in which the line is used by the various operators in rotation. It is interesting to note that this distributor type of multiplex utilizes the frequency-range of the line as efficiently as would a single printing telegraph channel using the same dotting speed, and more efficiently than does the carrier multiplex method. By the distributor method each operator utilizes the full frequency-range of the line during the time allotted to him and there is no time wasted in separating the channels from each other. In the carrier multiplex, on the other hand, while each operator uses the line for the full time it is available, a part of the frequency-range is wasted in separating the channels because of the departure of physical filters from the ideal. Also both side-bands are generally transmitted in telegraphy, in which case a still greater line-frequency-range is required for the carrier method.

If the message is produced originally as a continuous time function, as in speech, the same method may be used by substituting for the tape a phonographic record. That here also the required line-frequency-range varies directly as the speed of reproduction and inversely as the time of reproduction is obvious when we consider an imperfectly defined wave as equivalent to a succession of finite steps or a perfectly defined wave as a succession of infinitesimal steps. From the steady state viewpoint, all of the component frequencies are altered in the ratio of the reproducing and recording velocities, and hence the range which they occupy is altered in the same ratio.

Thus we see that for all forms of communication which are carried on by means of magnitude-time functions an upper limit to the amount of information which may be transmitted is set by the sum for the various available lines of the product of the line-frequency-range of each by the time during which it is available for use.

Application to Picture Transmission

However, if in order to utilize fully the line-frequency-range we introduce the process of recording, our message no longer exists throughout its transmission as a magnitude-*time* function, but becomes a magnitude-*space* function. Also in the case of picture transmission the information to be transmitted exists originally as a magnitude-space function. We may, of course, regard either a phonograph record or a picture as a secondary symbol, and say that the information transmitted consists of the sender's selection of a particular record

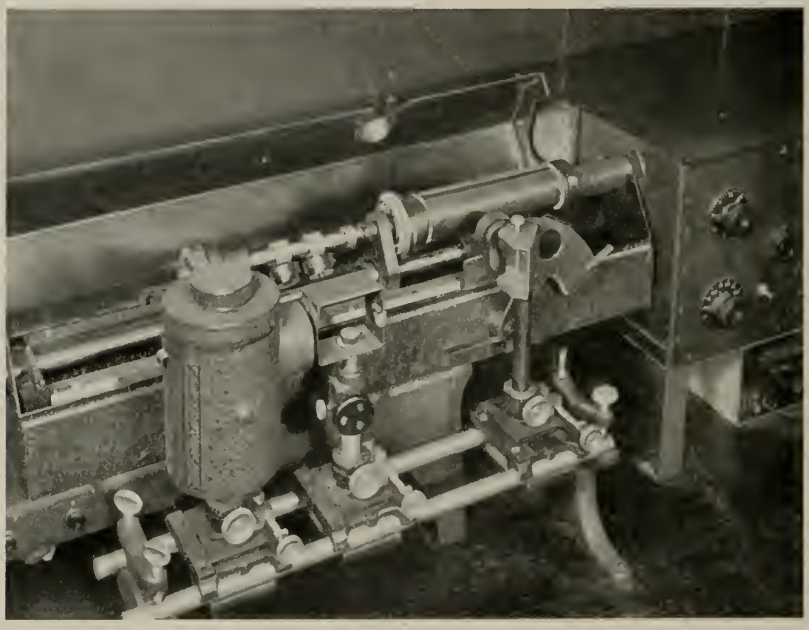


Fig. 6

or picture to which he desires to call the attention of the receiver. The information involved in such a selection is then measured by the logarithm of the number of different records or pictures which he might have selected. The problem then is to analyze the magnitude-space function which constitutes the secondary symbol into a sequence of primary symbols. This may be done in a manner similar to that already employed for magnitude-time functions.

The case of a phonograph record is directly analogous to those already considered in that the magnitude is a function of the distance along a single line. This distance is therefore analogous to time and

the information content may be found exactly as it would be from the pressure-time curve of the air vibration. In a picture, on the other hand, two dimensions are involved. We may, however, reduce this to a single dimension by dividing the area into a succession of strips of uniform width, as is done by the scanning aperture which is used in the electrical transmission of pictures. Figure 6 shows this scanning mechanism. The picture is mounted on a revolving cylinder which at each revolution is advanced by a spiral screw by the width of the desired strip. This scanning operation is equivalent to making an arbitrary number of selections in a direction at right angles to the strips. The number of these determines the degree of resolution in that direction. If the resolution is to be the same in both directions, we may consider the magnitude-distance function along the strip to be made up of the same number of selections per unit length. The total number of primary selections will then be equal to the number of elementary squares into which the picture is thus divided. These elementary areas differ from each other in their average intensity. The number of different intensities which may be correctly distinguished from each other in each elementary area of the reproduced picture represents the number of primary symbols available at each selection. Hence the total information content of the picture is given by the number of elementary areas times the logarithm of the number of distinguishable intensities.

In an actual picture the intensity as a function of distance along what we may call the line of scanning is a definite continuous function of the distance, but if there is any blurring of the picture as reproduced this function loses some of its definiteness. This blurring may be thought of as a form of intersymbol interference, since the intensity at one point in the distorted picture depends upon the original intensity at neighboring points. The similarity of this type of distortion to the intersymbol interference occurring in magnitude-time functions as a result of energy storage suggests that the picture distortion may also be treated on a steady state basis. We may think of the magnitude-distance function representing the picture as being analyzed into sustained components in each of which the intensity is a sinusoidal function of the distance. We may visualize such a single component in terms of the mechanism employed for recording and reproducing speech by means of a motion picture film. The intensity of the light transmitted by the developed film varies along its length in accordance with the magnitude of the electric wave resulting from the speech sound. If the speech wave be replaced by a sustained alternating current, there will result on the film a sinusoidal variation

in intensity with distance. The distance between successive maxima, or the wave-length, will vary inversely with the frequency of the applied alternating current. Figure 7 shows such a record of a speech wave and of sinusoidal waves of two different frequencies. The variations are superposed on a uniform component so as to avoid the difficulty of negative light.

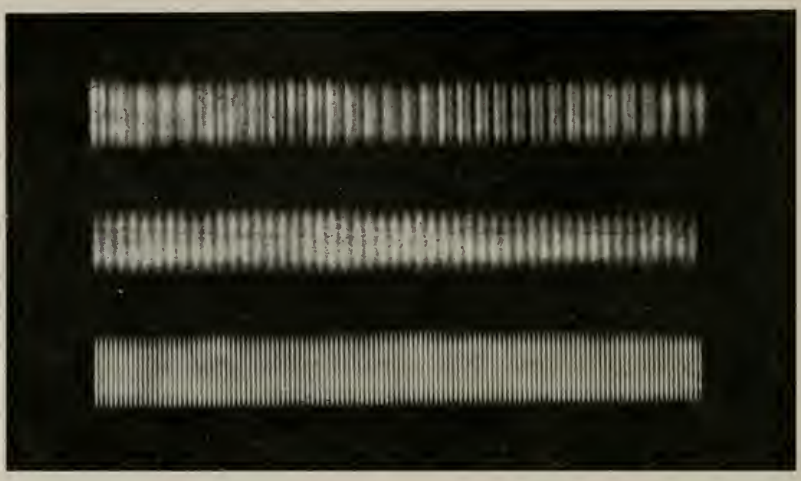


Fig. 7

The frequency of an alternating current is defined as the number of complete cycles which it executes in unit time. The analog of frequency in the corresponding alternating space wave is therefore the number of complete cycles or waves executed in unit distance. This is the reciprocal of the wave-length just as the frequency is the reciprocal of the period. Inasmuch as the term wave-number has been used by physicists to designate the reciprocal of wave-length, I shall use that term to designate the quantity corresponding to frequency in the steady state analysis of a magnitude-distance function. The distortion suffered by a picture in transmission may therefore be expressed in terms of the steady state amplitude and phase distortions as functions of wave number. Just as the transmission of a given amount of information requires a given product of frequency-range by time, so the preservation of a given amount of information in a picture requires a corresponding product of wave-number-range by distance. To illustrate, consider the effect of enlarging a picture without changing its detail or fineness of intensity discrimination. Suppose the enlargement to be made in two steps. In the first the

horizontal dimension is increased and the vertical dimension left unchanged. Let the scanning strips run in a horizontal direction. If we consider the magnitude-distance function representing the variation along any horizontal strip, the effect of the enlargement is to increase the wave-length of each steady state component in the ratio of the increase in linear dimension. The wave number of each component is therefore decreased in this ratio, and so the wave-number-range is also decreased in the same ratio. The product of the wave-number-range by the length of the strip remains constant, as does also the sum of the products for all of the strips, that is, for the entire picture. The second step consists in increasing the vertical dimensions with the horizontal dimensions fixed. By considering the scanning strips as running vertically in this case it follows at once that the product of wave-number-range by distance remains constant during this operation also.

Since the information transmitted is measured by the product of frequency-range by time when it is in electrical form and by the product of wave-number-range by distance when it is in graphic form, we should expect that when a record such as a picture or phonographic record is converted into an electric current, or vice versa, the corresponding products for the two should be equal regardless of the velocity of reproduction. That this is true may be easily shown. Let v be the velocity with which the recorder or reproducer is moved relative to the record. Let the wave-number-range of the record extend between the limits w_1 and w_2. If we consider any one component of the distance function which has a wave-length λ, the time required for the reproducer to traverse a complete cycle is λ/v, or $1/vw$. This is the period of the resulting component of the time wave, so the frequency f of the latter is the reciprocal of this, or vw. The frequency-range is therefore given by

$$f_2 - f_1 = v(w_2 - w_1). \tag{24}$$

If D is the length of the record, then the time required to reproduce it is

$$T = \frac{D}{v}, \tag{25}$$

from which

$$(f_2 - f_1)T = (w_2 - w_1)D. \tag{26}$$

This shows that the two products are numerically equal regardless of the velocity.

Application to Television

As our first illustration was drawn from one of the earliest forms of electrical communication, the submarine cable, it may be fitting to use as the last what is probably the newest form, namely, television. Here the information to be transmitted exists originally in the form of a magnitude which is a continuous function of both space and time. In order to determine what line facilities are needed to maintain a constant view of the distant scene we wish to determine the line-frequency-range required. This we know to be measured by the total information to be transmitted per unit time.

In the systems of television which have been most successful the method has been similar to that of the motion picture in that a succession of separate representations of the scene is placed before the observer and the persistence of vision is relied upon to convert the intermittent illumination into an apparently continuous variation with time. The first step in determining the required frequency-range is to determine the information content of a single one of the successive views of the scene. This may be determined exactly as for a still picture. The required degree of resolution into elementary areas and the required accuracy of reproduction of the intensity within each area determine an effective number of selections and a number of primary symbols available at each selection. These determine a minimum product of wave-number-range by distance. This in turn is equal to the product of line-frequency-range by time which must be available for the transmission of a single view of the scene. The time available is set by the fact that flicker becomes objectionable if the interval between successive pictures exceeds about one sixteenth of a second. Thus we have only to divide the product of wave-number-range by distance for a single picture by one sixteenth to obtain the line-frequency-range necessary to maintain a continuous view.

In the result just obtained an important factor is the interval necessary to prevent flicker. The tendency to flicker is, however, the result of the particular method of transmission. If it were practical to eliminate this factor, the required frequency-range might be somewhat different. We might, for example, imagine a system more like that of direct vision in which the magnitude-time function representing the intensity variation of each individual elementary area is transmitted over an independent line and used to produce a continuously varying illumination of the corresponding area of the reproduced scene. The frequency-range required on any one of these individual

lines would then be determined by the extent to which the intensity at any one instant could be permitted to be distorted by the inter-symbol interference from the light intensities at neighboring times; that is to say, the frequency-range necessary would depend upon a blurring in time analogous to the blurring in space which is used to set the wave-number-range for a single picture. It seems probable that the total frequency-range required would be somewhat less for such a system than for one in which flicker is a factor.

Conclusion

At the opening of this discussion I proposed to set up a quantitative measure for comparing the capacities of various systems to transmit information. This measure has been shown to be the product of the width of the frequency-range over which steady state alternating currents are transmitted with sensibly uniform efficiency and the time during which the system is available. While the most convenient method of operation does not always make the fullest use of the frequency-range of the line, as is the case in double side-band transmission, a comparison of the frequency-range actually used with that which would be required on the basis of the actual information content of the material transmitted gives an idea of what may be gained in the cost of lines by making sacrifices in the convenience or cost of terminal equipment. Finally the point of view developed is useful in that it provides a ready means of checking whether or not claims made for the transmission possibilities of a complicated system lie within the range of physical possibility. To do this we determine, for each message which the system is said to handle, the necessary product of frequency-range by time and add together these products for whatever messages are involved. Similarly for each line we take the product of its transmission frequency-range by the time it is used and add together these products. If this sum is less than the corresponding sum for the messages, we may say at once that the system is inoperative.

Carrier Systems on Long Distance Telephone Lines[1]

By H. A. AFFEL, C. S. DEMAREST and C. W. GREEN

SYNOPSIS: Two previous papers before the American Institute of Electrical Engineers discussed the activities of the Bell System in the development of multiplex telephone and telegraph systems using carrier current methods. The present paper describes developments which have resulted in improvements in the carrier telephone art during the past few years. A new, so-called type "C" system is described in detail, together with suitable repeaters and pilot channel apparatus for insuring the stability of operation; the line problems are considered and typical installations pictured. The growth of the application of carrier telephone systems and their increasingly important part in providing long distance telephone service on open-wire lines are shown.

INTRODUCTION

AT the 1921 Midwinter Convention of the American Institute of Electrical Engineers, Messrs. Colpitts and Blackwell presented a paper entitled "Carrier Current Telephony and Telegraphy." This described the development work of the Bell System and the resulting commercial types of multiplex telephone and telegraph systems using carrier current methods. The paper also gave a brief historical summary and included a theoretical discussion of the methods involved.

The carrier current art had at that time emerged from the laboratory to play its part in meeting the practical requirements of telephone service in the field. This step was made possible largely by two tools, now indispensable to the communication engineer, the thermionic tube and the wave filter.

In an ordinary telephone circuit, each frequency component in the voice of the speaker is transmitted by an electrical current of the same frequency. In most cases the electrical equipment of the circuit is not called upon to transmit frequencies above about 3,000 cycles per second. In carrier current operation, however, the voice-frequency currents are caused to modulate a high-frequency current which thus serves as a "carrier" for the message. In this way, an additional telephone channel is obtained, using frequencies entirely above those transmitted in connection with the ordinary voice frequency channel. By using other high frequencies, several additional messages may be transmitted simultaneously on the same pair of wires. Each channel occupies a certain range of high frequencies. For example, the words of one speaker may be conveyed by a channel employing frequencies from about 23,500 to about 26,000 cycles per

[1] Presented before the Summer Convention of the American Institute of Electrical Engineers, June 29, 1928.

second. At the receiving terminal the various incoming ranges of high-frequency currents are separated by electrical filters. Then by demodulation the original voice-frequency currents are produced again and are transmitted over voice-frequency circuits, the transmission over each channel thus reaching the proper listener. In this way a telephone line already carrying direct-current telegraph and voice-frequency telephone services may be multiplexed so as to provide additional telephone facilities. In a somewhat similar manner the high-frequency range may be used instead to transmit telegraph messages. In the present paper, carrier telephony alone is considered.

The Colpitts-Blackwell paper described two carrier telephone systems which had been developed up to that time, a four-channel "carrier suppressed" system (type "A"), and a three-channel "carrier transmitted" system (type "B"). The initial installation of these systems was made about 1918 on the long lines of the Bell System.

These earlier systems were effective in bringing about economies by avoiding the stringing of additional wire on many long pole lines, but there remained many opportunities for further improvement in performance and simplification of equipment. New problems arose to be solved in connection with the desire to operate the largest possible number of systems on the same pole line. The result has been the development of a substantially improved technique and a new system (the type "C") which not only has provided much improved performance over its predecessors but which has led to further economies because of reduced costs.

Carrier Telephone Growth in Bell System. Whereas the use of the early types of systems was justified in competition with the alternative of additional wire stringing only for distances exceeding 250 to 300 miles, the new system proves economical for distances considerably less. This fact has naturally stimulated the application of carrier telephony in the Bell System. This is shown by Figure 1, which indicates the growth of these systems in terms of channel mileage afforded by their use. It will be noted that the rate of growth of the systems has increased greatly in the last two or three years, a result of the availability of the improved system.

At the end of 1927 there were in operation about 130,000 channel miles. By the end of 1928 the figure is expected to be about 230,000. This figure does not, of course, represent a very large proportion of the total toll mileage of the Bell System, which includes many circuits less than 100 miles in length. It is sufficient, however, to indicate that the carrier telephone systems are a substantial factor in the provision for the growth of the longer haul facilities, where they naturally

provide the greatest economics. Their use is, of course, restricted to sections of the country in which open-wire construction is chiefly employed.[1] They have contributed toward lowering the cost of service and in making possible the toll rate reductions which have been put into effect within the past year or so.

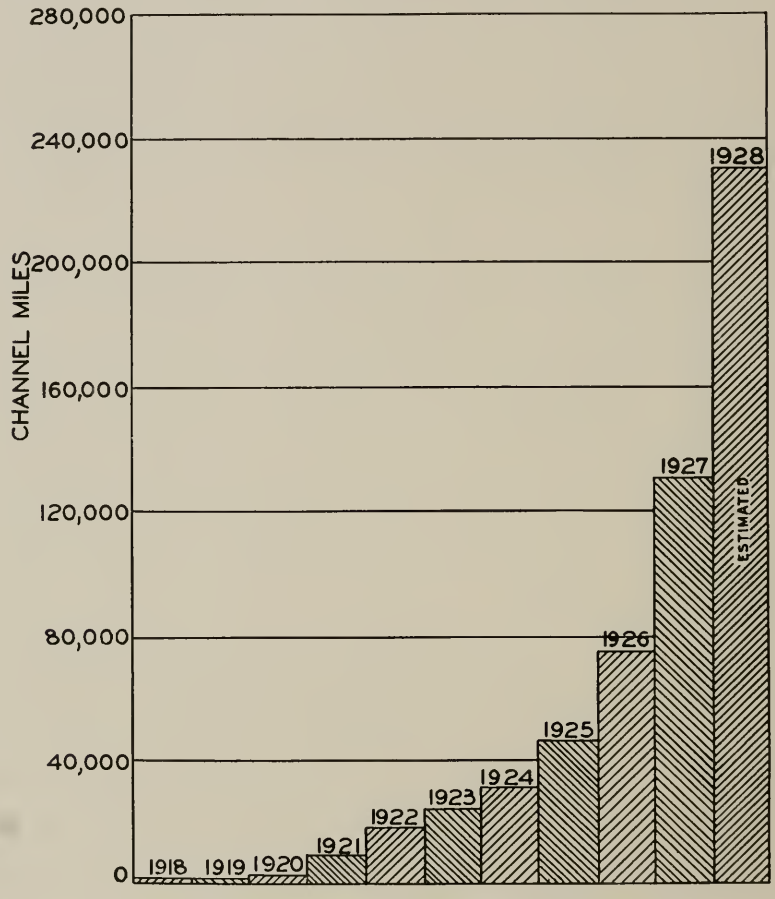


Figure 1—Growth of carrier telephony in Bell System

New System Replacing Older Types. The new type "C" system is essentially a long-haul, multi-channel system. It adds three high grade telephone circuits to the facilities normally afforded by a single pair of wires, and can be used over any distances likely to be encountered in the Bell System. Where repeaters are required they

[1] In localities having very heavy traffic requirements such as in the East, extensive use is made of toll cables.

are spaced at intervals of 150 to 300 miles depending upon particular transmission considerations. By means of a pilot channel, stability of transmission over the several carrier channels is assured, despite the relatively large inherent variations in high-frequency line transmission due to weather changes.

The service requirements which present themselves in the application of carrier methods are, of course, basically no different from those for commercial talking circuits obtained by other means. The problem is to establish a toll circuit between long distance offices which meets certain standards of transmission, including speech volume, stability and quality. The latter requires that there must be transmitted a certain band width of frequencies in the voice range. Furthermore, there must exist no appreciable load distortion effects. The circuit must also be relatively free from noise or crosstalk. A signaling system must be provided so that the operators at opposite terminals may call each other. In other respects the system must appear as a normal telephone circuit not distinguishable from an operating standpoint from the other circuits afforded by metallic wire connections. The apparatus installed in the telephone office must conform to certain physical standards of equipment, ruggedness, flexibility, etc. It must be capable of being maintained by trained office forces. Testing facilities must be provided, etc. It is believed that these objectives have been largely realized in the arrangements which are described in this paper.

The Type "C" System

The type "C" system embodies those major technical features which our experience with the older systems has indicated as most desirable. It is a carrier-suppressed, single sideband system, in which respect it is similar to the older type "A" system. However, it has been found possible to dispense with the equal frequency spacing of the channels which was characteristic of the type "A" system, and which involved the transmission of a synchronizing current between two terminals and the harmonic generation of higher frequencies from this synchronizing current. A simplification in apparatus has resulted. This non-harmonic arrangement of channels has further made possible a more efficient use of the frequency spectrum by the fact that the channel bands at lower frequencies can be squeezed together more closely than those of the higher frequencies where the band filters are less efficient due to decreasing ratio of band width to frequency.

The type "C" system requires for each modulator an oscillator as a source of carrier supply. Moreover, since a synchronizing current is not employed at the receiving terminal of the channel, an oscillator

of the same frequency is required for "demodulation." Advances in the art of designing vacuum tube oscillators of great frequency stability have made it possible to insure that these oscillators, which may be hundreds of miles apart, remain sufficiently close together in frequency so that no noticeable impairment in quality of transmission results.

In the matter of the frequency allocation of the channel bands, the type "C" system possesses one of the essential features of the older type "B" system, that is, the use of different carrier frequencies for transmission in opposite directions. Comparative experience with the type "A" system which, by means of high-frequency line and network balance, employed the same frequency band for the opposite directional paths of the channel led to the conclusion that the systems which avoided the high-frequency balance requirement were most desirable. Also the problem of intermediate repeater amplification is simplified where the opposite directional frequencies are thus separated and grouped. Furthermore, the crosstalk problem between different systems on the same pole line is greatly simplified for reasons which will be discussed later, and a greater total number of channels may usually be obtained on the same pole line.

The single sideband transmission employed reduces by about one half the frequency band that would otherwise be required for each channel. The carrier is not transmitted, as the presence in the system of carrier currents of the large magnitude required for a "carrier transmitted" system not only requires greater amplifier load capacity at the repeaters, but may increase the possibility of troublesome crosstalk and noise interference. The selectivity requirements of the band filters would also become more severe to keep the carrier of one channel out of the other channels in the system.

A Complete System. The simplified layout of a complete system is shown on Figure 2. It will be noted that it includes apparatus at a terminal, a line circuit, a repeater station, a second line circuit and apparatus at a second terminal. Obviously, the total line length between terminals may be extended by the use of a greater number of repeaters.

At each end there are the terminations of the three carrier channels 1, 2 and 3, and the regular voice circuit 4. These terminations appear, of course, at the long distance switchboard in the same office or in a different office from the carrier terminal. When a subscriber is connected to one of the terminations, for example, No. 1, speech currents pass through the three-winding hybrid coil, thence into the modulator circuit where they are caused to modulate high-frequency

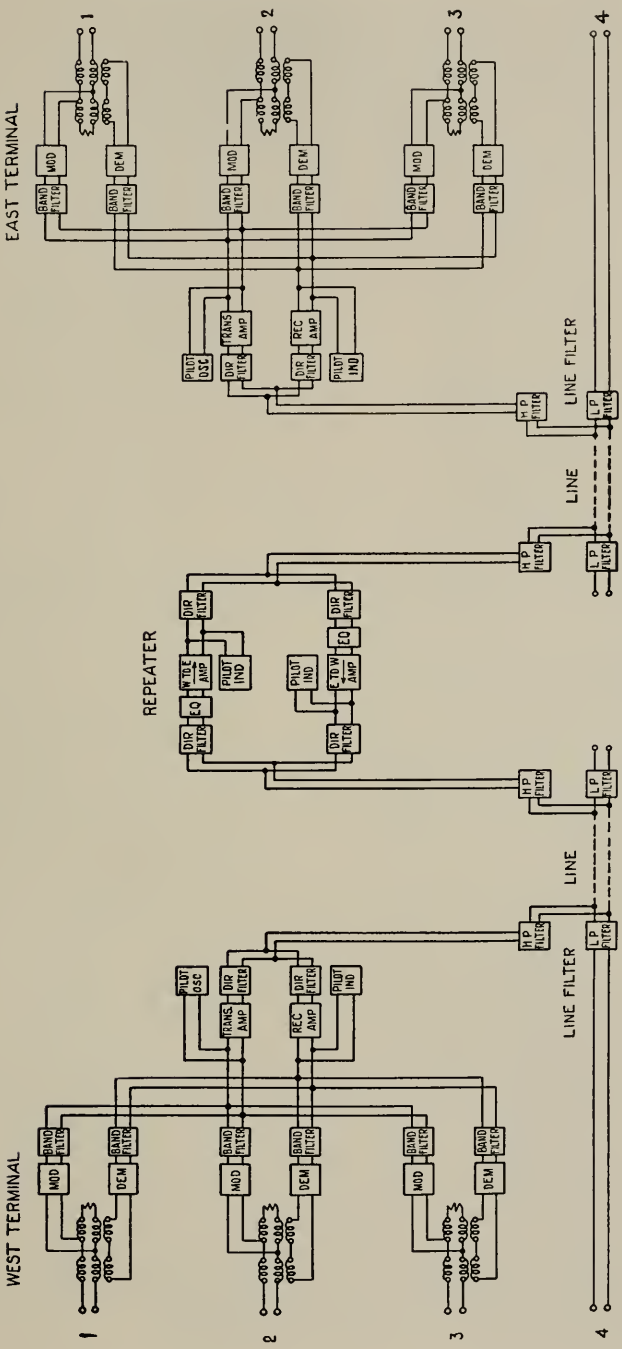


Figure 2—A complete carrier system schematic

carrier current. The resultant modulated bands [2] of frequencies pass through a band filter allowing only the desired band to pass to the transmitting amplifier, thence this band passes through a so-called directional filter and a high-pass filter to the line circuit. The high-pass filter last referred to, in association with its complementary low-pass filter, forms a so-called line filter set whereby the regular voice range currents are separated from the higher frequency carrier current at both terminal and repeater offices.

The other two carrier channels function similarly, and the several modulation bands of carrier frequencies join the first channel in passing through the common amplifier and directional filter circuit to the line. At the repeater point the group of bands comprising the three channels passes through the high-pass line filter circuit, thence through a directional filter and line equalizer to the amplifier circuit and outward through the directional and line filter circuit to the next line section. At the farther terminal the combined carrier currents pass through the directional filter and are again amplified in the receiving amplifier. At the output of the amplifier the different carrier channel bands of frequencies are selected one from another by the band filters, thence they pass to the demodulator circuit, are demodulated to their original form and then pass from the output connection of the hybrid coil to their respective terminations.

Circuit Arrangements at Terminals. Figure 3 shows diagrammatically in somewhat greater detail the terminal of the type "C" system. The modulator circuit consists of a two-tube "push-pull" grid-bias vacuum tube circuit in which the carrier frequency is balanced out. A separate oscillator tube circuit of exceptional frequency stability supplies the carrier. The frequency allocation requires the transmission of only the upper or lower sideband frequencies, and the band filter at the output selects the desired band, rejecting the other products of modulation as well as the amplified voice frequencies which are incidentally transmitted through the modulator unit. This sideband current in conjunction with the corresponding currents of the other two sidebands of the outgoing channels passes through the common amplifier. This is a two-stage vacuum tube unit having four tubes in the output circuit arranged in parallel push-pull connection to insure the required load carrying capacity.

The circuit then leads through a directional filter of either low-pass or high-pass type which distinguishes between the band groups of

[2] For a discussion of modulation see E. H. Colpitts and O. B. Blackwell, "Carrier Current Telephony and Telegraphy," *A. I. E. E. Transactions*, V. 40, 1921, pp. 205–300; R. V. L. Hartley, "Relation of Carrier and Side Bands in Radio Transmission," *Bell System Tech. Jl.*, V. 2, April 1923, pp. 90–112.

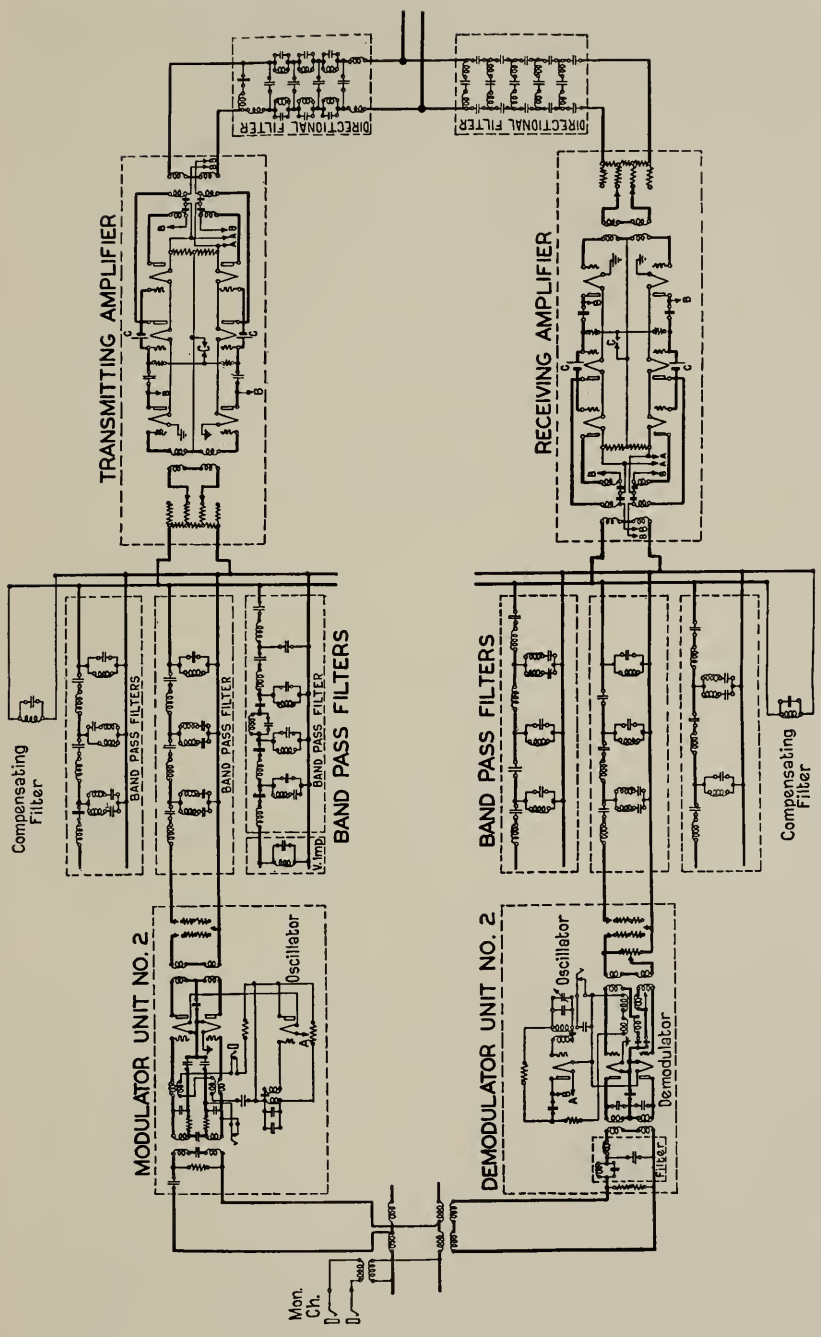


Figure 3—Schematic diagram of type "C" terminal circuit

the opposite directions of transmission as required by the allocation of frequencies. The amplified currents pass through the high-pass filter of the line filter set and thence to the line circuit.

In receiving, the sideband frequencies, after separation from the voice currents by the line filter set, pass through the directional filter and an amplifier similar to that used at the transmitting terminal. While the power output required at the receiving amplifier is usually small as compared to that required at the transmitting amplifier, the same unit is used for the two positions to provide flexibility in the adjustments of the receiving gains of the separate channels and for the purpose of economy in production. The different channel currents in the output of the amplifier are selected by the respective receiving band filters and thence pass into the demodulator circuits. In the demodulators the voice frequencies are derived by the modulation of the sideband currents with a carrier frequency supplied by a local oscillator whose frequency is adjusted accurately to agree with that of the corresponding transmitting modulator at the farther terminal. This important problem of synchronization of oscillators is further discussed later in the paper. It is, of course, obvious that if the carrier frequencies of the modulator and the corresponding demodulator of the same channel are not in sufficiently close agreement there will be a serious distortion of the speech currents received over the channel.

The output of the demodulator circuit includes a low-pass filter for suppressing the unwanted components of demodulation, and the circuit thence leads to the channel terminal through the hybrid coil. The function of the latter is to provide a two-wire termination of the channel and it prevents the output currents of the demodulator from reaching in any substantial magnitude the input of the modulator circuit, thus setting up a regenerative action which might result in "singing."

It may be noted that the circuit normally provides for a transmission "gain" or amplification of energy from the switchboard termination to the high-frequency line circuit of approximately 20 TU[3] corresponding to a current or voltage amplification of 10 to 1. In the receiving direction a gain of the same order of magnitude is also available. Of course, the exact amount utilized in a particular case depends on the line attenuation and the desired overall equivalent of the circuit. It is usually desirable at the transmitting terminal to maintain the level at the maximum possible for the system. The

[3] R. V. L. Hartley, "The Transmission Unit," *Electrical Communication*, V. 3, No. 1, July 1924, pp. 34–42. W. H. Martin, "Transmission Unit and Telephone Transmission Reference Systems," *A. I. E. E. Jl.*, V. 43, No. 6, June 1924, pp. 504–507, *Bell System Tech. Jl.*, V. 3, July 1924, pp. 400–408.

overall transmission afforded by a carrier system may be noted by the curve on Figure 4, which shows the relative speech frequency transmission characteristics of a typical channel. Where the carrier

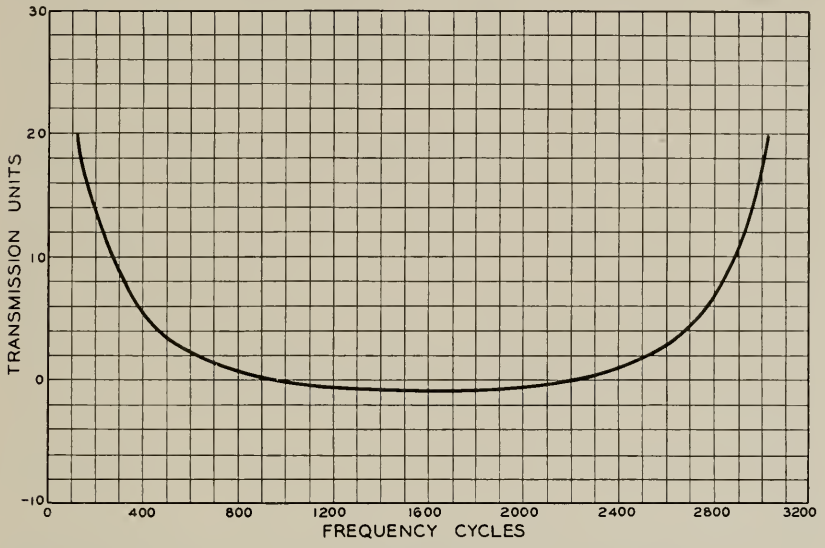


Figure 4—Representative overall transmission-frequency characteristic—type "C" carrier telephone system

channel is employed for terminal to terminal business the overall equivalent at 1,000 cycles is ordinarily adjusted to about 10 TU. The channels not infrequently form sections of much longer overall circuits, being connected to cable or perhaps open-wire circuits, in which case it is rather common to adjust the carrier section to a zero equivalent or even a gain of several TU.

Line Considerations. The passage of the carrier currents from the terminal apparatus over the line circuit which serves to connect the two terminals, or a terminal and repeater station, gives rise to several problems: the line loss or attenuation, the stability of transmission, the possibilities of crosstalk from other carrier systems on the same pole line and interference from currents from external sources. These factors must be considered not only in connection with the arrangement of the wires themselves but also in conjunction with the design of the terminal apparatus, repeaters, etc., so that satisfactory overall speech transmission may result.

As was brought out in the Colpitts-Blackwell paper, the line attenuation at the high frequencies is in accord with the recognized transmission theory. Because of skin effect in the wires and rising

losses in the insulators the attenuation increases steadily with frequency. Unfortunately the losses at the insulators are not constant and they increase greatly with the presence of moisture. This brings about an increase in attenuation in rainy weather. Fog, sleet and wet snow may greatly increase these attenuation changes. There is also a lesser source of variation due to temperature change and its effect on wire resistance.

If care is not observed, the carrier currents may be interfered with on the line circuits by crosstalk from other carrier systems and by miscellaneous currents which enter the circuit by induction from the outside. These latter manifest themselves as noise in the carrier channels. This makes it essential to use only the metallic circuit, i.e., two wires well balanced to ground for transmitting the carrier currents. The balance to ground must be maintained at a high degree by frequent transpositions in the wires. Even with these precautions unavoidable residual unbalances may permit a certain amount of interference to appear. The final remedy is to insure that the relations between the circuit length and the apparatus gains are properly considered in order that the speech currents may have ample margin above the noise currents at all points in the circuit.

In the matter of crosstalk between systems closely adjacent on the same line the situation is alleviated by providing two frequency allocations. (See Figure 5.) These are "staggered" with respect to

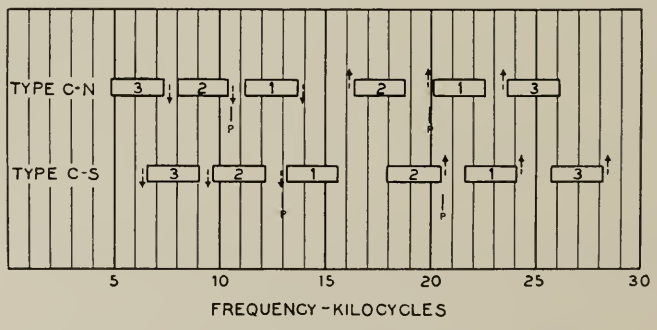


Figure 5—Frequency allocations of type "C" system

each other, so that a system installed on one pair using the so-called "N" frequency allocation has less crosstalk to and from a system installed and operating on an adjacent pair and using the so-called "S" frequency allocation than would be the case if both systems employed the same allocation. The maximum upper frequency required is raised only slightly by this arrangement.

Repeaters. Repeaters must be employed when the distance exceeds

that for which terminal transmitting apparatus is effective in maintaining the transmission level well above the line noise. The function of the repeater is, therefore, to amplify the carrier currents so that they pass on to the succeeding line section at a magnitude comparable to that sent out from the terminals. Obviously, the design of the repeater with respect to its gain and level carrying capacity, etc., presents a wide range of possibilities depending on the distance of transmission, frequency, etc.

It has been found most practical to install the repeaters along the route at approximately the spacing of the voice-frequency repeaters on the same wires. This means a spacing of from 150 to 300 miles, and occasionally slightly over 300. To have in the same office both voice-frequency and carrier repeaters reduces the equipment, simplifies the maintenance problem, and makes it possible to use the same sources of power supply. The gain and the load carrying capacity are, therefore, determined by this spacing, the gain being controlled by the attenuation loss between the repeaters, and the load carrying capacity by the output level desired because of noise considerations.

The higher attenuation of the line in the carrier range of frequencies means that the carrier repeaters must have a maximum gain of approximately four times that of the voice repeaters operated on the same wires. Whereas gains of the order of 8 to 15 TU may be readily supplied by voice repeaters using balance and so-called "two-wire" operation, the 30 to 45 TU gain required by the carrier repeaters necessitates non-balanced or "four-wire" operation or its equivalent, by using different frequencies in opposite directions and directional filters for the prevention of "singing."

Figure 6 is a schematic diagram of the circuits comprising a typical repeater station including loading, compositing apparatus and line filters. After passing through the high-pass line filter the carrier currents arrive at the high and low group directional filters which distinguish between the oppositely directed currents. These filters are substantially the same as those used for similar purposes at the terminal stations.

It is, of course, required in the design of the directional filters that in each direction the filters must pass a frequency band sufficient to transmit properly the three carrier channel bands. In addition to this the filters must present a loss outside of the transmission band which is sufficient to prevent the two-way amplifier circuit from "singing." This means that considering the closed loop circuit of the two amplifiers and the four directional filters the attenuation in this loop must be considerably greater than the sum of the gains or

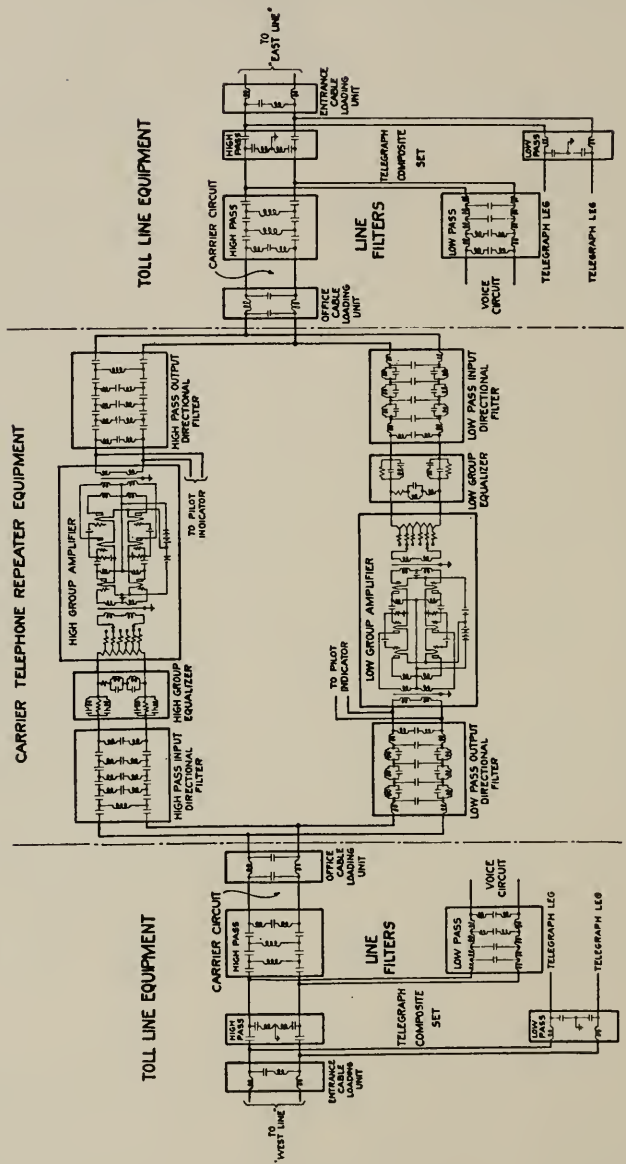


Figure 6—General schematic of carrier repeater circuit with associated line equipment

amplification of the two amplifiers. There are also other requirements which these filters must meet which are discussed later.

The amplifiers are the same as used for group amplification purposes at the terminals. Each consists of a two-stage reactance-coupled vacuum tube circuit having four tubes in parallel push-pull connection in the output circuit. The carrying capacity of this amplifier with the standard plate voltages is about one watt in the output, and the overall amplification or gain including incidental filter losses is about 30 TU. Where gains greater than 30 TU are necessary in the higher frequency group provision is made for the addition of an amplifier stage ahead of the unit shown, which adds approximately 15 TU gain. At the same time provision is made for the addition of greater directional filter selectivity.

An important feature of the repeater circuit is the equalizer which is connected ahead of the amplifier. Because the line circuit attenuation varies with frequency and is greatest at the higher frequencies it is necessary that the amplification introduced at a repeater point be varied with frequency. The amplification introduced by the amplifier unit itself is substantially uniform with frequency. The equalizer network, however, by introducing a loss which is a minimum at the highest frequency of transmission and which increases for the lower frequencies makes the overall repeater amplification a function of frequency and in general proportional to the line attenuation which it is designed to overcome.

A typical overall gain characteristic of the repeater is shown in Figure 7. The adjustment of the exact amount of gain desired at any time is made by the potentiometer at the input of the amplifier.

Pilot Channel. As noted previously, the attenuation of open-wire circuits of substantial length is affected by weather conditions. This makes it necessary to make occasional gain adjustments throughout the system. The extent of these adjustments is determined by means of the pilot channel, which provides a visual indication of the transmission levels of the carrier system in both directions of transmission without interfering with the speech currents over the channels themselves. It is, in effect, a separate constant frequency carrier channel allocated between certain speech channels in each transmission group.

The operation of the pilot is relatively simple. At each repeater point and receiving terminal there appears a meter for registering the output level of the amplifier. The pointer of the meter is expected normally to rest on the zero or normal level layout of the system. If a change in the attenuation of the line circuit causes a departure in the transmission level, the meter reading shows a corresponding

"up" or "down" indication and by adjustments of the repeater or terminal amplifier potentiometers the level may be returned to normal. An alarm circuit is furthermore provided at the receiving terminal

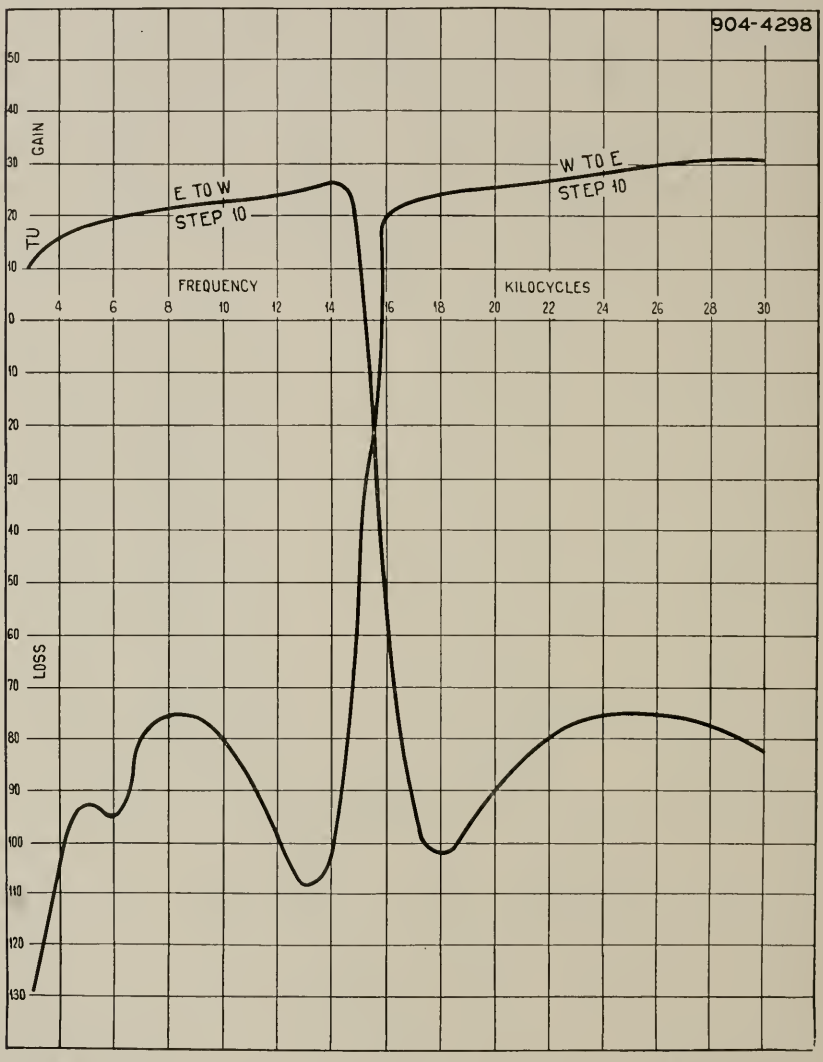


Figure 7—Overall transmission characteristics of carrier telephone repeater station

so that when the level has departed by more than a predetermined amount, say $\pm$ 1.5 TU, from the desired normal, the operating attendant is called in to make the adjustment.

A high-frequency current of constant amplitude is transmitted

from each end, and the meter indications are measurements of this current at the output of repeater amplifiers, and at the receiving terminal amplifiers (see Figure 2). A separate pilot frequency is utilized for each direction of transmission. Because no communication is carried on over this pilot carrier current, the band provided is extremely narrow, and no appreciable portion of the frequency spectrum is sacrificed.

The frequency selected for the pilot channel must coordinate with the other carrier system frequencies. The two frequency allocations of the type "C" system require different pilot channel frequencies, because their speech channels occupy different frequency bands. The apparatus has, therefore, been made so that the frequency of the pilot current can be adjusted to any value desired in the carrier range. The frequency selected for a given system may be determined by local conditions of crosstalk or interference, although in general the preferable location is between the channel bands as noted in Figure 5. The amount of current which is used is limited by its interfering effect into adjacent channels or into other carrier systems on the same line, and it is ordinarily of a low value, of the order of 2 to 6 milliamperes on the line.

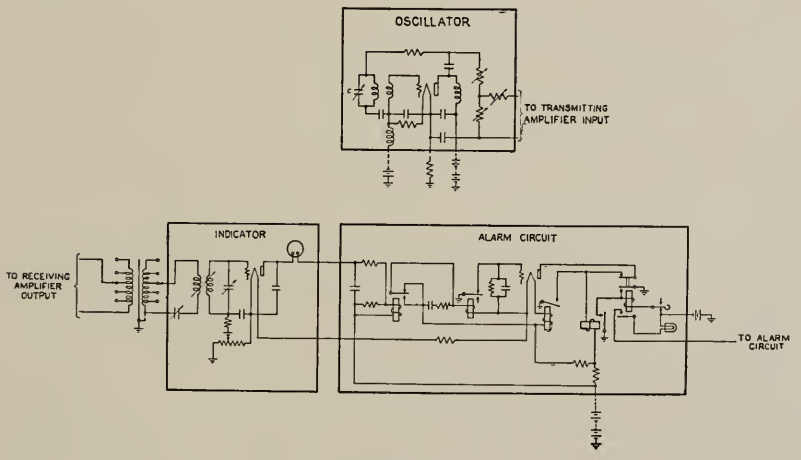


Figure 8—Schematic of pilot channel circuits. (The alarm circuit is used with terminals only)

Figure 8 shows schematically the principal features of the terminal pilot-channel circuit as a whole. The oscillator at each transmitting terminal which produces the pilot current is connected to the carrier circuit at the input to the transmitting amplifier, in parallel with the band filters. This current is amplified with the speech currents and

transmitted through the directional filter to the line. The attenuated pilot and sideband currents pass from the first section of the line into the receiving directional filter of the first repeater and enter the amplifier. The pilot channel indicator circuit is bridged across the output of the amplifier, and is tuned to discriminate very sharply against all but current of the pilot frequency. This circuit has a high impedance relative to the line, so that only a very small percentage of the pilot current is drawn from the line at a repeater point. The remainder is transmitted through the outgoing directional filter and over the subsequent section of the line.

That portion of the pilot current which enters the indicator circuit is amplified and rectified in the vacuum tube detector, and the output current is read on a d.-c. milliammeter. As stated above, this meter is calibrated to read in TU above and below a mid-scale position which represents a normal transmission level to which the system is initially adjusted.

Entering the receiving terminal of the carrier system, the pilot and speech currents pass through the directional filter and are amplified. As at the repeater, the pilot indicator circuit is bridged across the output of the amplifier. At this terminal, in addition to showing level, the output of the indicator actuates an alarm circuit which operates when the transmission level at this point varies from normal for a set interval of time by more than a prescribed amount. This delay action in the operation of the alarm provides selectivity against slight interference into the pilot channel from currents on the other channels of the system and thereby insures that the alarm indicates a definite level change.

The pilot channel thus insures that the high-frequency portion of the system is continuously checked with the exception of the individual channel band filters and modulator and demodulator units. These, however, are particularly stable in operation and require no unusual attention in maintenance. Of course, the overall check is made at only the pilot frequency in each direction. Variations of line equivalent caused by weather changes increase in magnitude with frequency. Therefore, corrections must be made in the gain relations of the individual channels whenever these weather changes are great. Fortunately the corrections follow a fairly definite relation with variations of pilot level and are ordinarily made by the terminal attendants on the channel potentiometers controlling the demodulator gain by reference to a table. This table shows the relations between the required gain changes at the three channel frequencies in terms of changes at the pilot frequency.

The type of oscillator is essentially the same as that used in the type "C" carrier systems for producing the carrier frequencies. It is controlled by condensers which include an adjustable air condenser for tuning to the particular frequency desired.

Two indicators are located at the repeater, one for each direction of transmission. Each indicator circuit consists of a vacuum tube rectifier operating from coupled tuned circuits into a d.-c. milliammeter having a special scale calibrated in transmission units. The filament and plate currents and bias potentials are obtained from the standard 130-volt battery. The advantage of using the same battery for the several functions is that it makes possible the stabilization of the rectifier output with power variations. An adjustable grid bias voltage is obtained from the negative drop of the filament circuit with an opposing 3-volt dry cell battery connected in series. With this arrangement normal variations in the 130-volt source cause only a negligible change in the indicator meter readings.

At the receiving terminal, in addition to the indicator circuit which is the same as at the repeater, an alarm circuit is provided as noted above. A sensitive marginal relay is connected in series with the indicator meter. When this relay operates, it starts the delay circuit by removing ground from the grid condensers of the alarm tube. The leakage through the grid resistances then causes the condenser potential, which is the grid potential of an auxiliary rectifier tube operating from the same power source, to decrease slowly, resulting eventually in a rise in the current of the plate circuit of the alarm rectifier tube. If the marginal relay remains operated for a given length of time, the alarm tube plate current will rise to a value necessary to operate the alarm relays. For shorter periods of operation, the normal highly negative grid potential of the rectifier tube is restored and no alarm is operated. The timing of the delay circuit is adjusted by the values of the grid leak resistances and condensers. A delay of about 15 seconds is usually employed, which effectually prevents false operation due to occasional transients such as speech interference. The adjustment of the contacts on the alarm relay is ordinarily such as to cause an alarm to be given at limits of $\pm$ 1.5 TU variation.

General Transmission Considerations

Lines. The typical open-wire telephone line consists of a number of 10-foot crossarms spaced two feet apart on poles whose height varies from 30 feet upward depending on local conditions. The poles are spaced at an average interval of 130 feet. Each crossarm carries 10 wires. The wires are normally spaced at 12-inch intervals, except

in the case of the so-called pole-pairs which straddle the pole and whose wires are about 18 inches apart. (See Figure 9.) The construction includes pins and glass insulators for supporting the wires.

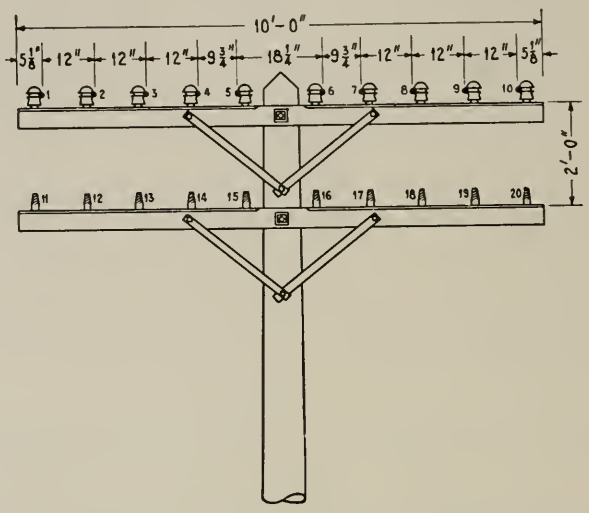


Figure 9—Showing arrangement of wires on telephone pole line

There are three gauges of wire in common use in the telephone plant, having diameters of 104, 128 and 165 mils,* respectively. The largest gauge, 165-mil pairs naturally afford the lowest attenuation and have been generally used in connection with the application of the longer systems. The pairs of this sized conductor are, however, now fairly well used up for carrier purposes and new installations are being made more often on the smaller diameter circuits.

Typical attenuation curves for the three gauges of wire and the extremes of weather conditions are given in Figure 10. It will be noted that the wet weather attenuation may be as much as 40 per cent higher than the dry weather attenuation. Also, these variations are greater at the higher frequencies.

It is interesting in this connection to consider the effect of the possible variation in a practical case. Take, for example, a 165-mil pair 200 miles long with a carrier channel frequency at 25 kilocycles. This means a total attenuation of 20 TU in dry weather and 29 TU in extremely wet weather, a variation of 9 TU or a current ratio of about 3 to 1. In the case of a still longer line these possible variations present rather startling figures. For example, in a 1,000-mile circuit the variation would be five times the above or 45 TU, which would

*The term "mil" as here used is equivalent to 0.001 inch.

mean that if the circuit were set up to have a proper volume of transmission in dry weather and rain occurred over the whole line it would cause the speech at the receiving end to drop to but 1/180 of the

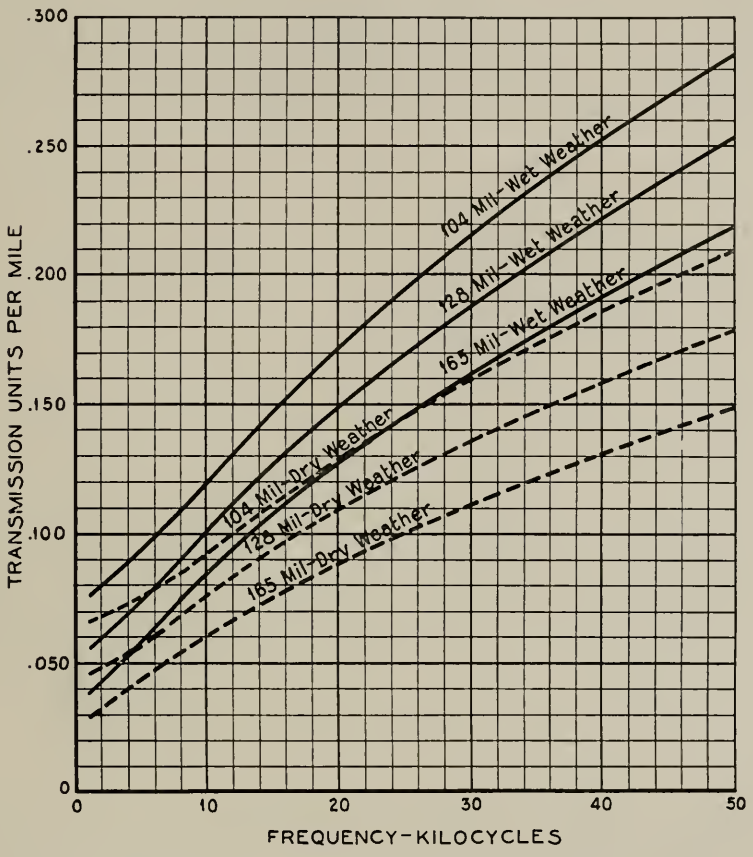


Figure 10—Attenuation curves for open-wire lines of different gauges at high frequencies

desired volume if the proper readjustments of gain at the repeaters and terminals were not made. Fortunately, these line variations occur gradually, at least in the case of the longer lines.

In connection with most carrier installations measurements are made [4] of line characteristics prior to the installation of the apparatus.

[4] Reference, "High-Frequency Measurements of Communication Lines," by H. A. Affel and J. T. O'Leary, *A. I. E. E. Transactions*, V. 44, 1927, pp. 504–513.

An interesting picture is presented in Figure 11 which shows the attenuation variations with time on a particular line (about 110 miles in length) during the period in which a storm arose to cause the attenuation to increase. Later, when the insulators dried, the corresponding drop in attenuation was that shown. From these variations it is quite obvious that means such as afforded by the pilot channel are needed to insure that the talking circuits provided by the carrier channels remain at substantially constant volume.

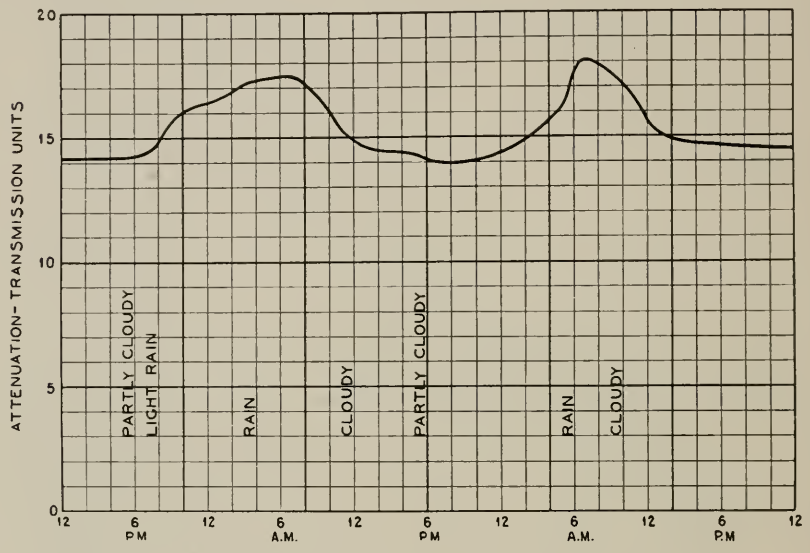


Figure 11—Variations in attenuation of a particular open-wire circuit

In addition to the improvement in stability effected by the use of pilot channel apparatus, substantial advances have been made in the design and application of special types of line insulators in which the high-frequency losses, particularly in wet weather, have been appreciably reduced, resulting in still further improvement in stability. The attenuation data given above are for the lines equipped with the older standard types of telephone insulators, which are still employed on the majority of circuits in the telephone plant. However, the newer types of improved insulators are now being applied and their use makes it possible to reduce the wet to dry weather attenuation variation by a factor of about 3 to 1 and to reduce the absolute value of attenuation at the higher frequencies by as much as 25 per cent. Further information describing the development

work which has made possible these improved insulators will be made available at a later date.

While the circuits employed for the transmission of carrier telephone systems as noted above are largely of open-wire construction, where these circuits pass through the more populated districts of the country it is frequently necessary to insert sections of cable. The smaller closely spaced wires of cables make the problem of attenuation at high frequencies more serious, even where the cables are relatively short, say a mile or so in length. Typical attenuation curves of non-loaded cable pairs are shown in Figure 12.

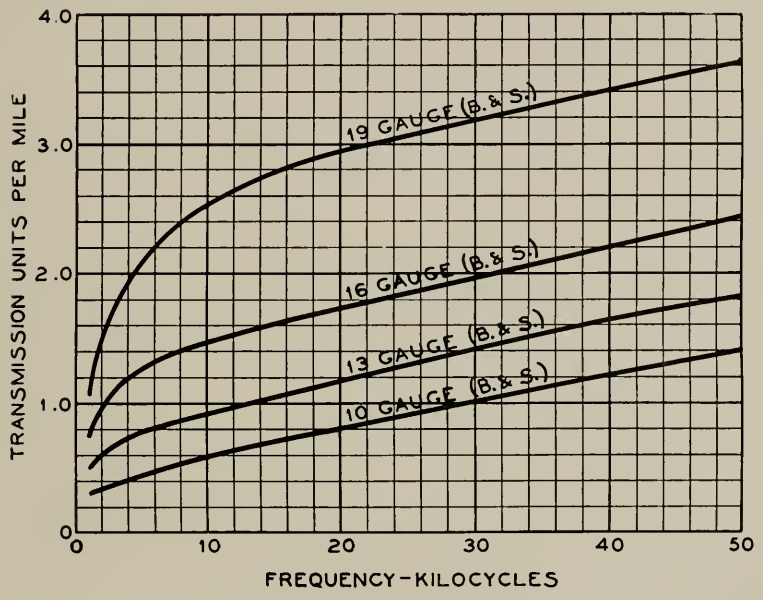


Figure 12—Attenuation of non-loaded cable circuits

This situation has led to the development of a special type of cable loading which permits making a substantial reduction in the attenuation for the higher frequencies and which also makes the characteristic impedance of the cable circuit more closely simulate that of the open-wire circuit so that the reflection effects discussed in detail later are thus greatly reduced. This is important, for, whereas the open-wire circuit characteristic impedance varies from 600 to 700 ohms, the non-loaded cable impedance is of the order of 130 to 150 ohms and the reflection losses and also certain resultant crosstalk effects as discussed later are, therefore, very substantial for even short lengths of non-

loaded cable. The present standard types of carrier cable loading systems [5] provide for the use of loading coils spaced at intervals of approximately 930 feet. When loaded, the cable circuits have a characteristic impedance closely approximating the open-wire impedance over the frequency range used in carrier transmission. This same carrier loading also greatly improves the characteristics of the voice circuit. The high-frequency attenuation is reduced to approximately one half the non-loaded condition. A special type of cable loading is also available for use in improving the transmission characteristics of office cable and wiring and very short intermediate and entrance cable.

External Interference. The carrier channels are unusually free from noise due to extraneous induced currents. However, this is the result of attention to this factor in the design of the apparatus and in laying out the installations rather than anything inherent in the high-frequency feature as such. Our experience has indicated that it is possible, if care is not taken, to have interference from the following external sources:

a. Harmonics of power frequencies.
b. Irregular frequencies produced by abnormal power line actions, such as arcing insulators, charging lightning arresters of certain types, electric railways, series street lighting, etc.
c. Power line carrier systems.
d. Powerful transoceanic radio transmitters.
e. Lightning and other atmospheric disturbances.

In the matter of harmonics of the power line frequencies, the source of their generation normally limits them to very low magnitudes in the high-frequency range which has been employed for carrier systems on telephone lines. In this respect the carrier systems are, in general, affected to a lesser extent than the normal telephone circuits in the voice range. In the latter case, the power circuit harmonics frequently present serious interference problems because the harmonics in the power circuits are substantially greater at the lower frequencies.

Under particular conditions, however, such as, for example, in connection with a series street lighting system operated with individual series transformers or auto-transformers, where a burned-out lamp causes the saturation of the transformer magnetic circuit, induced harmonics of considerable magnitude, up to 30,000 cycles and over, have been measured in the carrier telephone circuits. Under the same

[5] Thomas Shaw and Wm. Fondiller, "Development and Application of Loading for Telephone Circuits," *Bell System Tech. Jl.*, April 1926, pp. 221–281.

conditions, however, much larger harmonics are present in the voice-frequency range, so that the induction in the normal telephone circuit is much more severe than the carrier circuit.

A much more severe source of carrier interference has been found to result from the abnormal actions of power line circuits in which arcing phenomena occur. Interference of this sort has been noted and traced to such sources as arcing insulators, tree leaks, pantograph and trolley collector sparking, charging lightning arresters, unusual commutator or slip ring sparking, switching, etc. In the early days of operation of carrier systems, interference of this type formed a not uncommon source of disturbance. The situation was remedied in some cases by cooperation with the power companies concerned. On the whole, this source of interference has been greatly reduced in the past few years.

On occasions the carrier telephone systems have been interfered with by power line carrier systems operating on near-by power lines. Considering the widespread use of power line carrier telephone systems and the fact that they normally involve a transmitting power many times that of the systems described in this paper, this would, no doubt, be a more common source of difficulty if it were not a fact that such power systems adjacent to the telephone systems are operated well above the frequency range of the telephone line carrier systems.

Energy picked up from the high-power transoceanic radio telegraph stations, transmitting at frequencies in the carrier range, is an occasional source of interference, particularly in the east where carrier systems are located relatively close to the radio stations. The open-wire telephone lines act as long-wave antennæ and intercept the radio energy. This, of course, enters initially on the longitudinal wire circuit to ground. Due to residual line unbalances, some energy is, however, unavoidably passed on to the metallic circuits on which the carrier systems are operated, and enters the speech channel in the form of a tone or note similar to a heterodyne signal at a radio telegraph receiver.

Lightning and general static disturbances form a substantial part of the background noise which is found on all carrier lines. Its general magnitude is ordinarily small, except under certain conditions such as the case of near-by storms.

Transmission Levels. In the design and laying out of type "C" installations, the transmission level of a system is ordinarily not permitted to fall below a certain figure, which under particular circumstances might be about − 25 TU, with respect to the trans-

mitting terminal. A transmission level diagram will serve to explain this limitation.

Let it be assumed that it is desired to effect carrier transmission using a type C-N system between points A and B, 240 miles apart on 165-mil conductors. The highest frequency channel is normally considered, which in this case would be 26 kilocycles. The total attenuation of the line at this frequency, as determined from the line attenuation data already presented, would be 35 TU for wet weather conditions of operation. A level diagram would accordingly picture the situation as noted in Figure 13. At point A sufficient

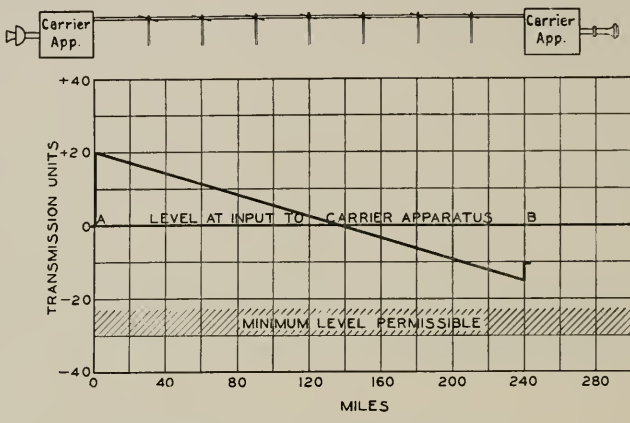


Figure 13—Transmission level diagram

transmitting gain would be provided by the equipment to bring the sending level to + 20 TU. The line attenuation in connection with transmission over the 240-mile circuit at point B would bring the level to − 15 TU. In order to obtain an overall talking circuit of, say, 10 TU., it would be necessary to operate with a receiving gain of 5 TU. It will be noted that in this particular layout the minimum line level is well above the limit set above. In fact, computations would indicate that the line circuit might be extended to the total length of about 300 miles, before the level limits would be exceeded. On longer lines, however, involving may repeater sections, the level limits are raised because of the cumulative effect of noise entering the circuit from a greater number of sources.

The line circuit illustrated is of the simplest type and in a practical case involving sections of intermediate and terminal cable construction the attenuation would be considerably greater and the effective geographical distance covered for a particular type of apparatus would, therefore, be less.

Crosstalk. Telephone circuits which are simultaneously operating in close proximity on a pole line are normally subject to crosstalk because of the mutual inductance and capacity relations between the wires. The problem which this presents in a pole line structure carrying many circuits requires careful consideration, even where the frequencies are no higher than the voice range. The problem is cared for by the application of transposition systems, i.e., arrangements whereby the effect of these relations between the circuits tends to be canceled out by transposing the wires constituting the two sides of a circuit in an orderly fashion. These transposition systems are carefully designed and the transpositions to be applied in each circuit specified.[6]

When using still higher frequencies for carrier purposes, this problem is correspondingly increased as the mutual relations tend to become greater at higher frequencies. The phase changes as the currents progress along the lines are more rapid for the higher frequencies. The design of the transposition system capable of permitting the simultaneous operation of a number of carrier systems on the same pole line is a difficult problem. The subject is one of great complexity and to give it complete consideration would require more space than is available here. It may be noted, however, that, by means of special transposition layouts installed in the circuits being used for carrier transmission, successful operation is being obtained with a large number of carrier systems on the same pole line, both telephone and telegraph. The locations of transpositions in circuits used for carrier transmission occur more frequently than in circuits restricted to operation at voice frequencies, in some cases as frequently as every other pole.

Several factors in the apparatus design have contributed to lessen the hardship imposed by the crosstalk problem:

1. The standardization of arrangements whereby the same frequencies are only employed in a given direction on systems on the same pole line.
2. The equalization of the transmission levels between paralleling systems.
3. The use of "staggered" frequency allocations for systems in closest proximity.
4. A careful consideration of impedance relations in the line circuits and apparatus.

[6] "The Design of Transpositions for Parallel Power and Telephone Circuits," H. S. Osborne, *A. I. E. E. Transactions*, V. 37, June 1918, pp. 897–936.

Frequency Directions. The importance of the use of a separate frequency for each direction of transmission may be considered by reference to Figure 14. If there are two paralleling telephone circuits

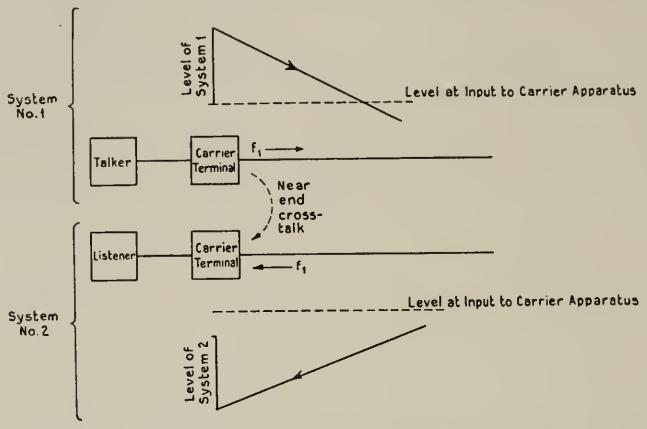


Figure 14—Diagram illustrating occurrence of near-end crosstalk between carrier systems employing the same frequency for opposite directions of transmission

employing frequencies (f_1) in the same range, and if there exists between the two circuits a certain amount of crosstalk, when there is a talker at the terminal of one system (No. 1) and a listener at the same terminal of the other system (No. 2), then the speech from the talker at the high level will enter directly into the sensitive receiving circuit of the listener. This is commonly called "near-end" crosstalk. In the case of a carrier circuit, the transmitting terminal would involve a certain amount of amplification. The receiving circuit would likewise, so that the net effect would be that the crosstalk between the two circuits would be amplified by the combined amount of gain or amplification present in the sending and receiving circuits. In telephone parlance it would be stated that this is a situation in which substantial level differences exist between the two circuits.

On the other hand, in the case of two adjacent carrier systems employing the same frequencies for the same direction of transmission, a crosstalk situation involving only "far-end" crosstalk would exist, as illustrated in Figure 15. This assumes that near-end crosstalk by reflection as discussed later has been eliminated. In this case the talker and the listener would be situated at opposite terminals of the paralleling circuits and the crosstalk, while being amplified like the near-end crosstalk by the total gain in the transmitting and receiving circuits, suffers the attenuation of the line circuit which more than offsets the amplification. This is, therefore, a very substantial factor in favor of the two-frequency method of operation.

At the carrier frequencies, it has been found impracticable to design transposition arrangements providing for systems where the same frequencies are transmitted in opposite directions. It has been found that, while the two-frequency operation may mean fewer two-way

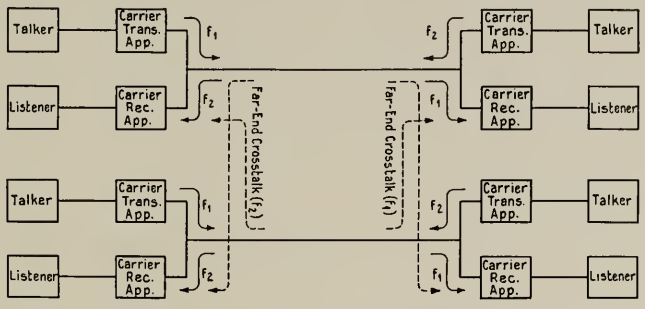


Figure 15—Diagram illustrating occurrence of far-end crosstalk only in carrier systems employing different frequencies for opposite directions of transmission

operating channels within the same frequency range on a single pair of wires than would be the case if the same frequency bands were provided for opposite directional transmission, the net result in the former case is to make it possible to obtain a greater number of channels on a pole line having many pairs of wires. The need for the directional coordination of frequencies has led to the general adoption of rules throughout the Bell System whereby the systems are all installed so that the low-frequency directional group of channels transmits east to west or north to south and the high-frequency directional group in the reverse direction, west to east or south to north.

Level Equalization. A situation involving an exaggeration of the crosstalk between two paralleling carrier systems may, of course, arise, even in the case of systems involving the transmission of the same frequency in the same direction for the two systems, if the transmission levels of the systems are not the same. If, for example, two systems operating between the same terminals are set up to have the same overall talking equivalent, and one system has a transmitting gain 10 TU higher than the other, the second system will have to have 10 TU greater receiving gain in order to provide the same overall equivalent. This would mean that this system would receive from the first system 10 TU higher crosstalk than if the levels of the two systems were alike. Efforts are, therefore, made in "lining up" the paralleling systems on a pole line so that as nearly as possible the same level relations are obtained for all systems, and the crosstalk tendencies are thus minimized.

Staggering of Frequency Bands. A substantial reduction of crosstalk is obtained through the staggering of an adjacent system frequency allocation as previously noted. Figure 5 shows the frequency allocations of the C-N and C-S systems. Because present standard types of telephone transmitters and receivers have response characteristics which exhibit the greatest sensitivity in the vicinity of 1,000 cycles, as the bands of two adjacent channels are shifted from an overlapping position, the crosstalk is appreciably reduced. In this case also the overlapping crosstalking points are always opposite sidebands and the intelligibility is completely lost even for the case of a substantial overlapping.

It is customary to install C-N and C-S systems on the two side circuits of a phantom group. The phantom group comprises four wires which are most closely associated electrically because they are employed not only to provide a telephone circuit on each pair of wires but a phantom telephone circuit each side of which is comprised of one pair of wires in parallel.

A typical arrangement of facilities afforded by one crossarm of the telephone line is illustrated by Figure 16. It will be noted that this

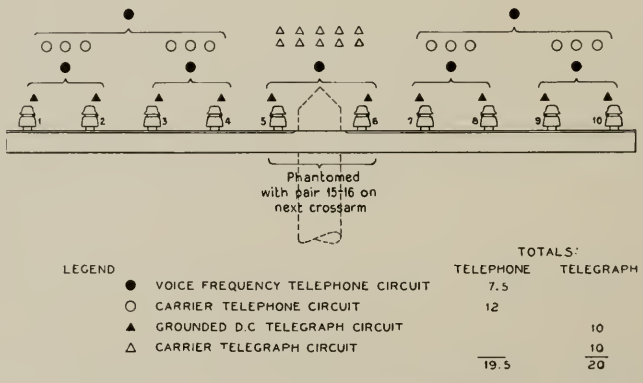


Figure 16—Arrangement of communication facilities on one crossarm

crossarm provides a total of twelve carrier telephone channels, five regular telephone circuits, two and one half phantom circuits, thus making a total of nineteen and one half telephone circuits. The telegraph facilities would total ten regular grounded d.-c. telegraph circuits, one for each wire, and ten carrier telegraph channels on the pole pair, thus affording a total of twenty duplex telegraph channels. This is, therefore, an average of approximately four telephone channels and four telegraph channels per pair of wires, which is obviously a fairly efficient use of the copper wire.

Impedance. It is found desirable in connection with communication circuits in general to match carefully the impedances of the various circuit and apparatus components if for no other reason than to insure the best transmission by keeping the reflection losses at a minimum. In connection with carrier systems the matter of crosstalk constitutes an additional important reason for doing this. As noted above, the crosstalk situation is simplified by the standardization of frequency arrangements by which only far-end crosstalk is normally received. This not only reduces the level differences at which crosstalk takes place as explained, but it simplifies the transposition design problem because near-end crosstalk is normally greater in magnitude than the far-end crosstalk. However, if the line circuit is irregular, i.e., if there are abrupt impedance differences in the circuit as it passes from point to point which bring about wave reflections, these may result in near-end crosstalk being reflected and appearing as far-end crosstalk, thus adding to the true far-end crosstalk and making it more difficult to keep within desirable limits. For this reason every effort is made in the layout of the carrier lines to avoid such reflection effects. This makes it desirable to load even relatively short cables including office cables and wiring. The apparatus terminal impedances are also carefully designed, so that their values simulate the characteristic impedance of the line circuits over which the systems are operated.

Overall Line Circuit. A situation sometimes occurs in a long carrier system where the line is made up of sections in which the wire pairs

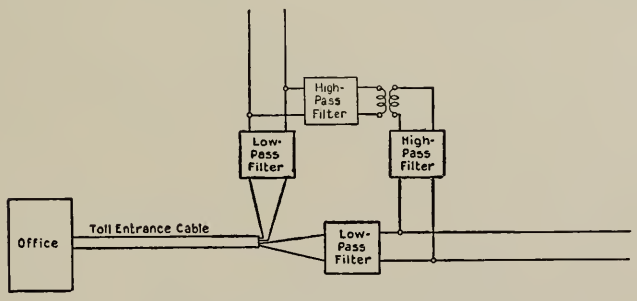


Figure 17—Schematic of high-pass transfer line filter circuit

occupy different pin positions in each section and the voice circuit on the pair in which the carrier system operates is terminated at different points or perhaps joins other lines. The use of line filter sets at the intermediate points makes this arrangement possible. Special transfer line filter sets have also been designed where it is desired to transfer the carrier currents from one pair of wires to another without affecting

the destination of the voice circuits and with a minimum impedance irregularity for either circuit. These line filters are sometimes mounted on poles, so that this transfer may take place where lines join at an outside point and where office equipment cannot be installed. A circuit arrangement illustrating the use of the pole-mounted high-pass transfer filter set is shown in Figure 17.

Equipment Problems and Typical Installations

The increasing use of carrier telephony as a substitute for line construction in providing toll facilities on long circuits has, like the development of toll cables, resulted in further increasing the proportion of the plant investment represented by the equipment within the offices. It has likewise required that a greater part of the maintenance effort involved in taking care of a given number of facilities be devoted to the equipment. These factors have made the design and arrangement of the carrier equipment matters of considerable importance. Recent developments in these respects have, therefore, been directed toward obtaining a high degree of adaptability of the carrier equipment to practical use in the telephone plant. Economies in design have also resulted which have been an important factor in extending the usefulness of the equipment.

The type "C" carrier telephone equipment is mounted on panels employing a uniform dimensional system in a manner similar to the other recent telephone developments. Arrangements have been devised so that in the future this mounting method will permit the desired close association between the carrier filters and other related apparatus in the lines in order to minimize high-frequency losses and impedance unbalances within the offices. Signaling arrangements flexibly adapted to present plant conditions have been provided.

The high frequencies and power levels used in carrier telephony and the frequency conversion functions of the system are the principal electrical factors which affect the arrangement and amount of equipment involved. The high frequencies necessitate careful wiring, shielding, and location of certain units with respect to others to avoid undesirable inductive and impedance effects. The modulation and demodulation processes and the high energy levels required necessitate the use of numerous vacuum tubes, with the consequent need of suitable sources of power.

Typical System Equipments. As noted previously, a long carrier telephone system involves equipment at a number of intermediate repeater stations in addition to that at the terminals. Figure 18

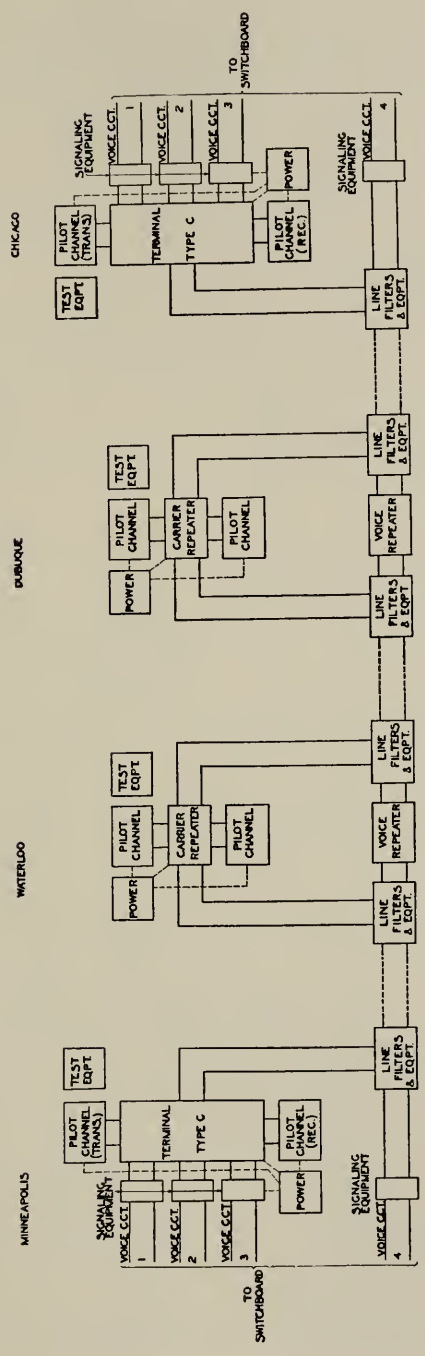


Figure 18—Elementary diagram showing principal equipment groups in typical type "C" carrier telephone system

shows the principal equipment groups involved in a typical long carrier system. The particular system illustrated is one of those between Minneapolis and Chicago, with repeater stations at Waterloo and Dubuque, Iowa. The carrier repeater equipment is ordinarily additional to voice-frequency repeater apparatus used in the wire line as mentioned above and is connected so that the high-frequency currents for the carrier pass around the voice-frequency repeater. The wires concerned are employed also for d.-c. telegraphy by the use of composite sets.

The principal groups of equipment involved in such a system include the carrier sending and receiving equipment and filters at the terminals, the repeater amplifiers and filters at the intermediate stations, and the line equipment and pilot channel equipment at all points. In addition, power supply equipment and testing equipment are required at all points, and voice-frequency and signaling apparatus at the terminals.

The total amount of equipment involved in a typical carrier telephone system shown in Figure 15, exclusive of the power supply, includes altogether about 188 panels assembled on racks equivalent to 14 bays [7] and occupying a total floor space, including aisle space, of about 84 square feet. If the three channels which the system ordinarily provides were obtained by regular wire circuits, the office equipment might amount altogether to about 36 panels and 1.7 bays, occupying about 10 square feet. Thus, in a typical case, about eight times as much office equipment, other than that for the power supply, might be required to furnish a given number of facilities by carrier telephony, in comparison with that needed for the equivalent number of ordinary wire circuits.

Terminal Station Installations. The principal equipment groups comprising a terminal of a type "C" system are indicated in Figure 19. A typical assembly showing a majority of these equipment groups is given in Figure 20. This does not include the signaling equipment, the pilot channel, or the power equipment. A rear view of this same assembly is shown in Figure 21.

Returning to Figure 20, the right-hand bay contains the apparatus comprising two channel terminals. The middle bay includes the third channel apparatus and the terminal transmitting and receiving amplifiers and directional filters which are mounted in the upper portion. The box-like units on both bays are the band filters and directional filters. On the right-hand bay the upper of the panels

[7] A bay consists of two channel or I beam uprights, ordinarily about 11½ feet high, and spaced so as to mount unit panels 19 inches wide and of varying height.

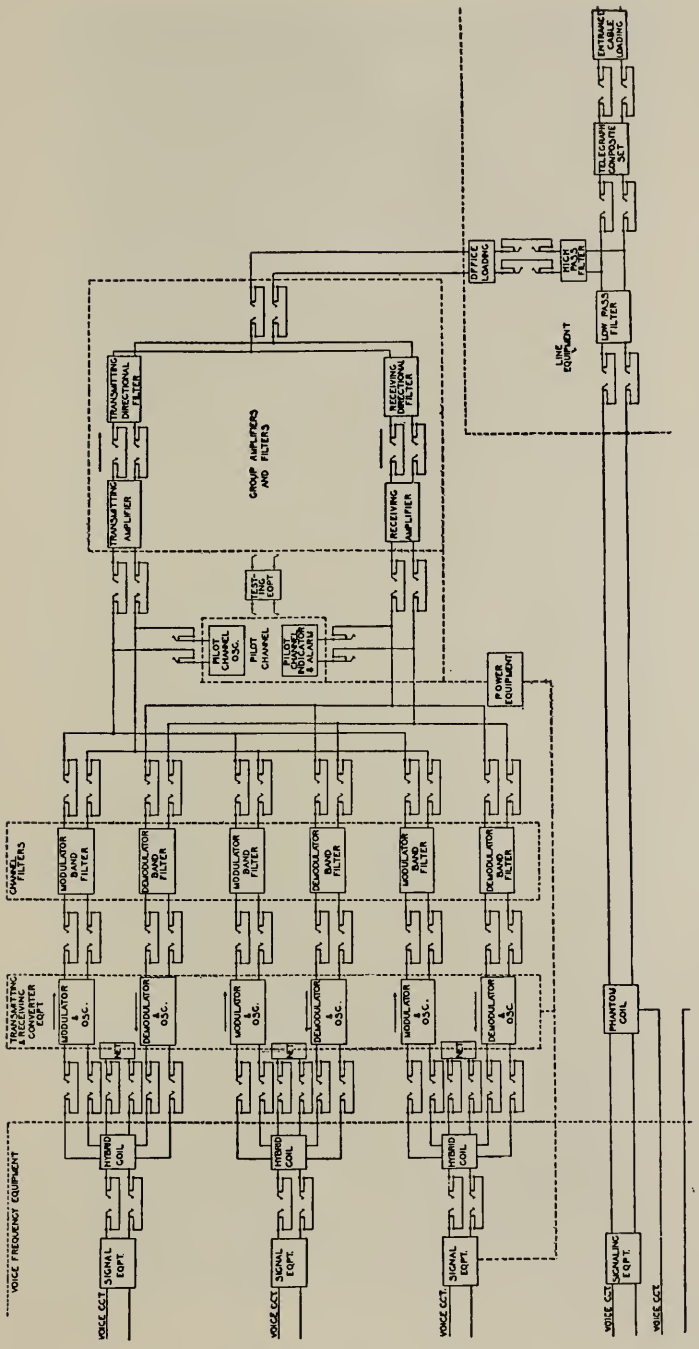


Figure 19—Schematic showing principal units comprising type "C" terminal equipment

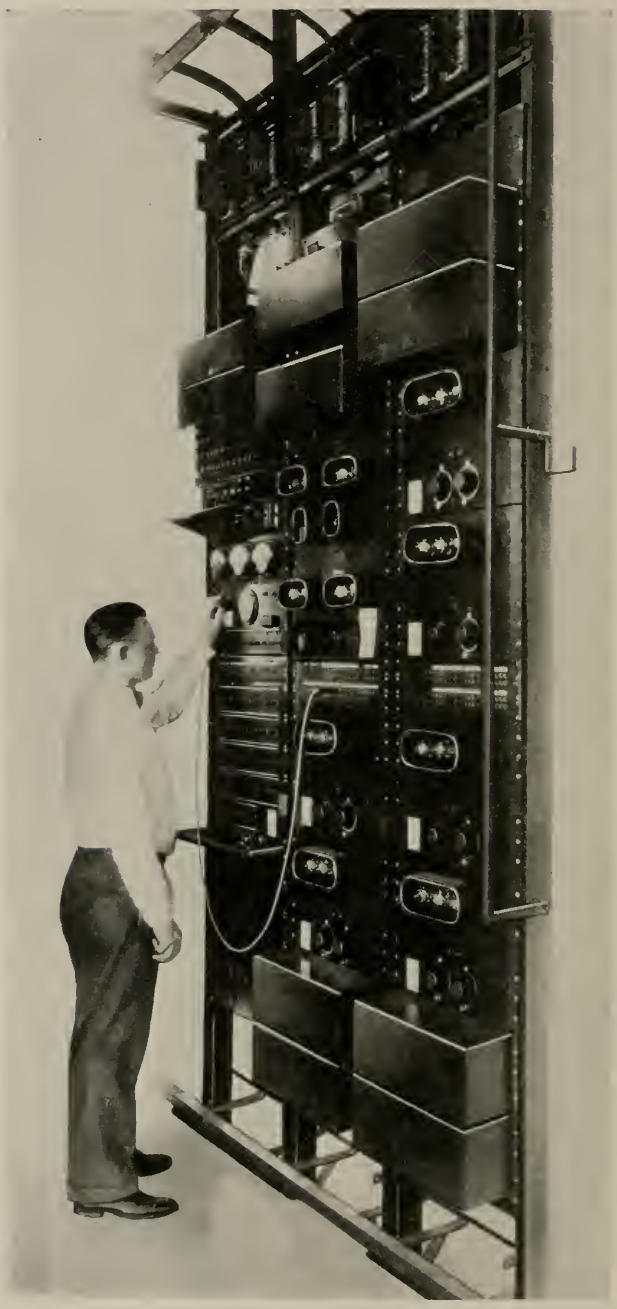


Figure 20—Type "C" carrier telephone terminal equipment, typical assembly of system. (Front)

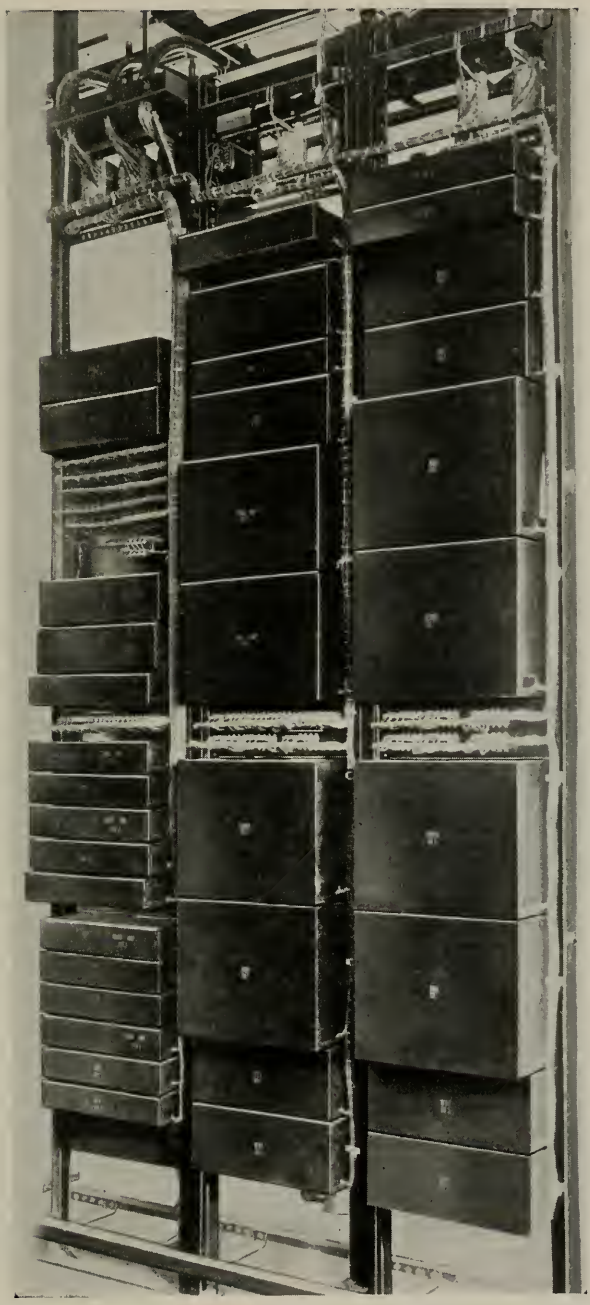


Figure 21—Type "C" carrier telephone terminal equipment, typical assembly of system. (Rear view)

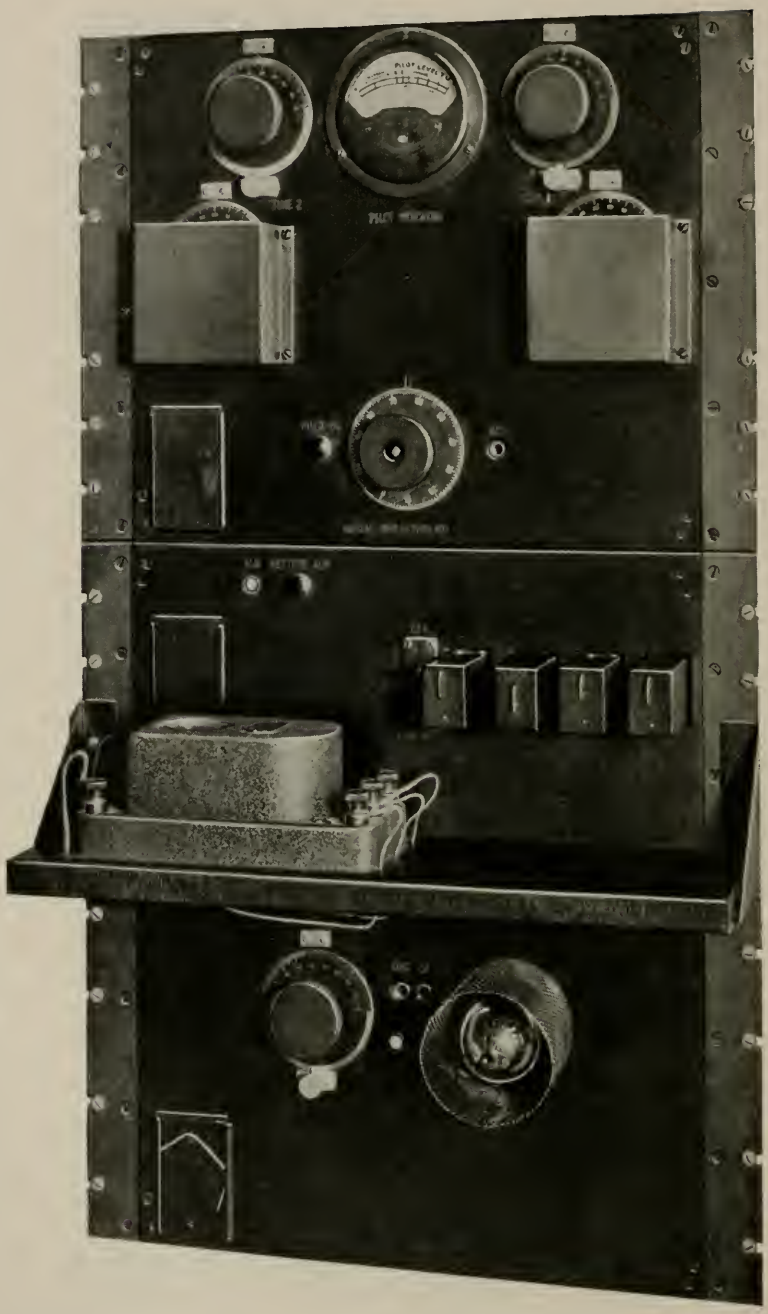


Figure 22—Typical assembly of pilot channel equipment in terminal installation

with three vacuum tubes is the modulator-oscillator panel of one channel. Below it is the demodulator-oscillator panel of the same channel. Below the latter and in the center of the bay is the jack mounting strip which makes it possible to disconnect, or switch for testing purposes, the various units of the complete equipment. The

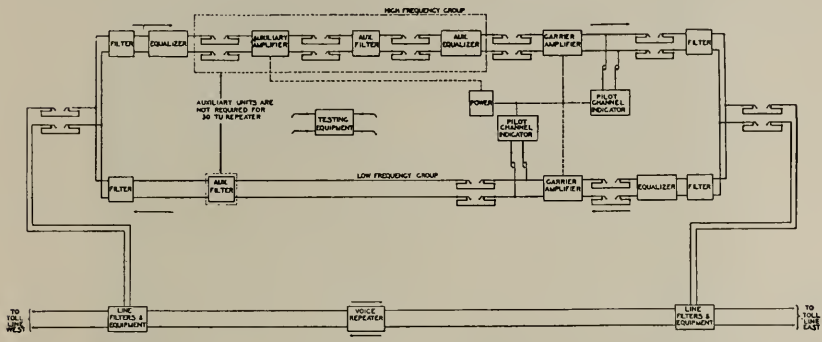


Figure 23—Schematic showing principal equipment units in carrier repeater circuit

demodulator-oscillator panel and modulator-oscillator panel, respectively, are the next two panels of the second channel. The terminal strips will be seen at the top of the bays. The metal shields surrounding the vacuum tubes are useful for mechanical protection only. The testing and power distribution equipment is located in the left-hand bay.

The pilot channel apparatus at the terminal station, which is employed in regulating the performance of the system to compensate for variations in the line equivalents, is assembled in a typical installation as shown in Figure 22. This apparatus may be located adjacent to the carrier terminal apparatus. The upper panel is the indicator unit with the indicator meter shown in the upper center of the panel. The panel immediately below this is the alarm panel with its voltmeter relay. On both of these panels the associated vacuum tubes are mounted in the rear. The lowest panel is the oscillator panel with its vacuum tube and frequency control.

Typical Repeater Station Installations. The equipment at each carrier repeater station consists mainly of the units indicated in Figure 23. It is seen from this figure that the principal items are the line filter equipment, the amplifiers, the equalizers, the directional filters, and the jacks provided for testing and patching the equipment. Pilot channel equipment, power supply equipment, and testing equipment are also included. The amplifier equipment in each carrier repeater,



Figure 24—Typical installation of carrier telephone repeater equipment. (Front view.)

other than the 15 TU auxiliary amplifier, is identical to the group amplifiers employed in each terminal station. The filters are also identical in type, excepting that twice as many directional filters are required at each repeater station as at each terminal.

A typical complete installation of two carrier repeaters with testing and battery supply circuits located in the bay at the left is shown in Figure 24. The bottom panel on the left-hand bay is a reserve amplifier. Above this panel are the filament rheostats, telegraph instruments, jack panels, key panels for controlling the power supply, a panel containing a thermocouple and meter for testing, meters for reading currents and voltages, and finally the alarm relays. These last are operated by failure in the plate current in the amplifier tubes, thereby indicating when a tube burns out, or failure of either A or B battery supply.

Each repeater bay in this case, Figure 24, contains an auxiliary amplifier to increase the gain in the high-frequency group. From

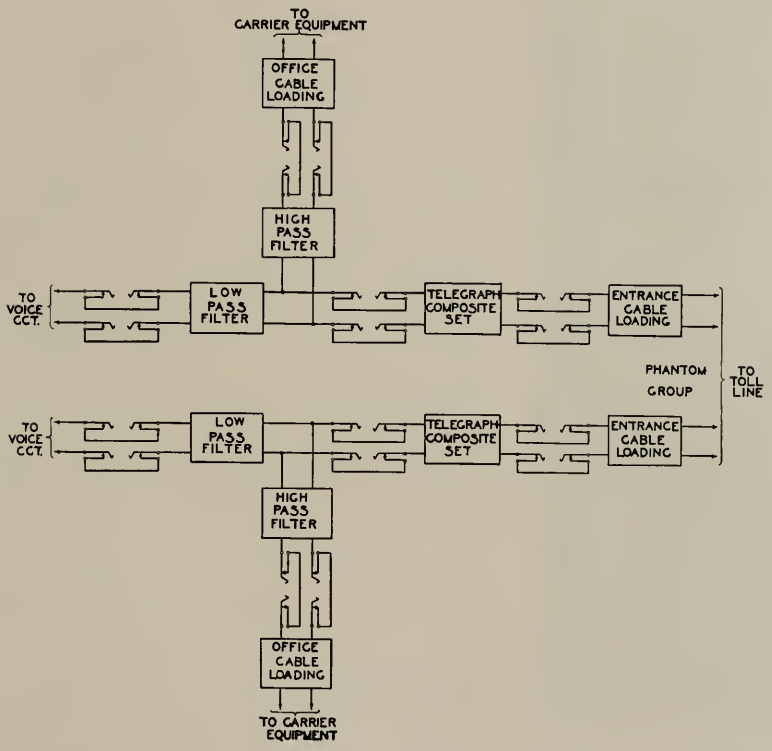


Figure 25—Schematic circuit showing line filter equipment for type "C" carrier telephone terminal. (Phantom group)

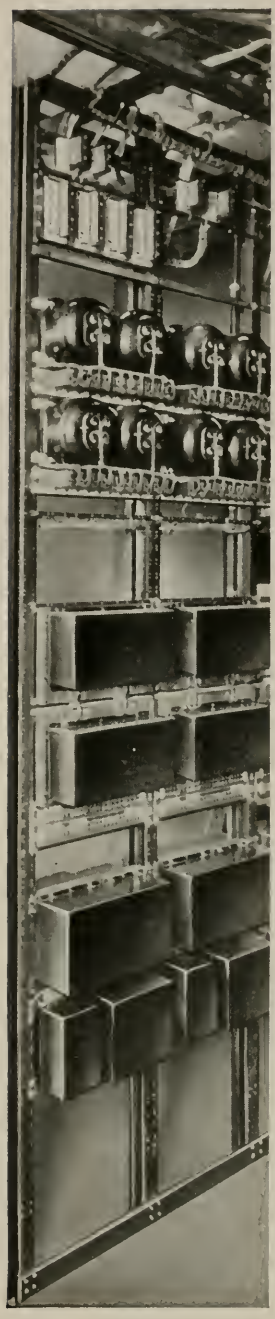


Figure 26—Arrangement of line equipment in one assembly for type "C" carrier telephone. (Front view)

top to bottom the panels are input filters, auxiliary filters, equalizers, auxiliary amplifier, two regular amplifiers separated by a jack panel, and output filters. The arrangement of filters and equalizers is chosen to minimize any tendency toward inductive feed-back effects between the output and input circuits of any one repeater as well as crosstalk between different repeaters on adjacent bays.

The pilot channel equipment at a carrier repeater station is similar to that employed at the terminal stations, excepting that the alarm and oscillator panels are not included. The alarm apparatus is omitted in this case, since it is not the practice to have the carrier repeater attendants readjust the carrier repeaters to take account of line changes, excepting when instructed to do so by the attendants at the terminal stations where the alarm apparatus is installed. The pilot channel equipment at each repeater station thus consists principally of an indicator panel associated with the transmission circuit in each direction, which is assembled with the other equipment as previously shown at the extreme right in Figure 24.

Line Equipment. Figure 25 shows the principal line equipment units which are closely associated in effecting connection between the high-frequency circuit of the carrier system and the voice-frequency line. This equipment, consisting of the line filters, composite sets, and entrance and office load coils, is mounted together in one assembly and located as near as practicable to the other carrier equipment. A method of assembly which is now under development is shown in Figure 26. The bay shown contains the line equipment for two phantom groups.

This compact method of assembling and wiring the line equipment reduces the amount of office cabling required for the carrier and,

therefore, reduces the possibility of inter-system crosstalk between carrier systems within the same office. The crosstalk requirements in an office may be more severe than on the pole lines because the level difference between circuits which operate on different pole lines terminating at the same office may be as much as 50 TU. As an aid in obtaining the required electrical separation all high-frequency wiring is reduced to a minimum by segregating and mounting together all line equipment associated with a single circuit. No high-frequency circuits appear at the toll testboard. The toll lines may be tested from the testboard by means of trunks between the testboard and the line equipment bays. All the carrier equipment is thoroughly shielded in such a manner that the separation between the equipment of any two systems is 120 to 135 TU.

The carrier line equipment at a carrier repeater station is generally similar to that at the terminal stations, as previously shown in Figure 26. At each repeater station, however, this equipment is provided in the lines in both directions. Two types of low-pass line filters are employed at the repeater station, one adapted to circuits in which both carrier and voice-frequency repeaters are used and the other, which is less commonly used, for circuits employing only carrier repeaters and where the voice circuit continues through without a repeater.

Voice-Frequency and Signaling Equipment. The general function of the voice-frequency terminating equipment is to associate, by means of a hybrid coil and network, the ordinary two-wire circuits in the tele-

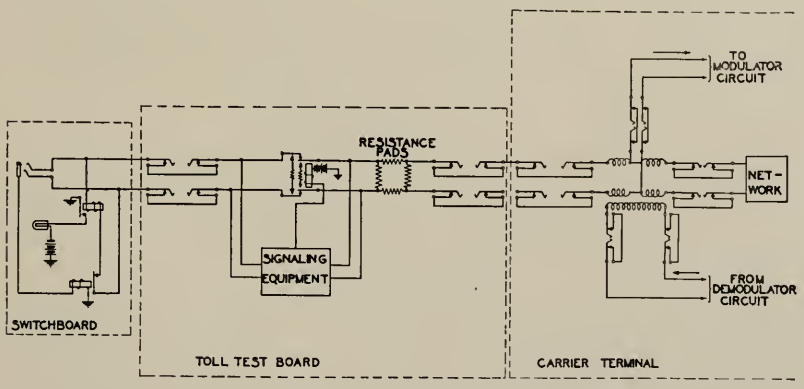


Figure 27—Schematic of voice-frequency circuit for type "C" carrier telephone system

phone switchboard, including both talking and signaling functions, with the sending and receiving branches of each of the different

carrier channels. The general arrangement of this equipment in the circuit is shown in further detail in Figure 27. This includes the signaling equipment and makes provision for the connection of a balancing network.

The use of a two-wire termination for the carrier system is necessary because ordinary telephone circuits at the switchboards, such as trunks, subscribers' lines, etc., are of the two-wire type and the cord circuits for interconnecting these are of this type. Hence, such a termination of the carrier system makes it possible to connect it to other circuits with the same apparatus and in the same manner as with ordinary telephone circuits.

The signaling apparatus consists of a 1,000-cycle ringer of the type which is employed on long voice-frequency lines. This is connected to the voice-frequency terminal of the carrier system in the same manner as to other voice circuits. The use of 1,000-cycle signaling with the carrier has been desirable in place of the more simple low-frequency signaling apparatus used on shorter lines, since frequencies less than 200 cycles are not efficiently transmitted.

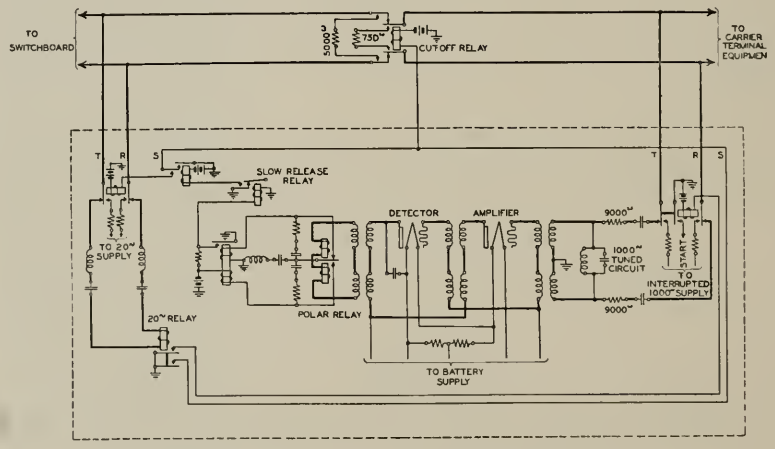


Figure 28—Schematic of 1,000-cycle ringing circuit

Figure 28 shows a simplified diagram of this type of ringer. The transmitted signaling currents as impressed upon the carrier channel are of 1,000-cycle frequency interrupted at a speed of 20 interruptions per second. This ringing current supply is obtained either from 1,000-cycle generators or vacuum tube oscillators. Such currents, while in the voice-frequency range and thus capable of being transmitted readily, form a signal of sufficiently distinctive character to permit separation from ordinary voice currents. Thus, practical

freedom from voice interference with the receiving apparatus is obtained, since this apparatus is designed to respond to very small currents of this character but to discriminate sharply against other currents.

Tests and Adjustments of Apparatus. For the purpose of testing and "patching" (i.e., interconnecting various equipment units in the carrier system), jacks are provided as previously shown in Figure 19. The equipment which is employed for testing and adjusting the carrier apparatus provides means for measuring gains and losses at the various frequencies encountered and includes a supply of testing current at these frequencies. The general arrangement of the testing apparatus provided for measuring gains and losses is shown in Figure 29. This is assembled with the other carrier equipment as previously shown in Figure 20. It consists chiefly of means for switching a known loss in the testing circuit, an attenuator, and a calibrated thermocouple type measuring instrument for determining the value of the current transmitted through this apparatus. The 1,000-cycle current is used for practically all testing of the terminal apparatus.

The testing equipment at a carrier repeater station is arranged in a manner similar to that at the terminal stations, as previously shown in a general way in Figure 29. It requires in addition, however, a

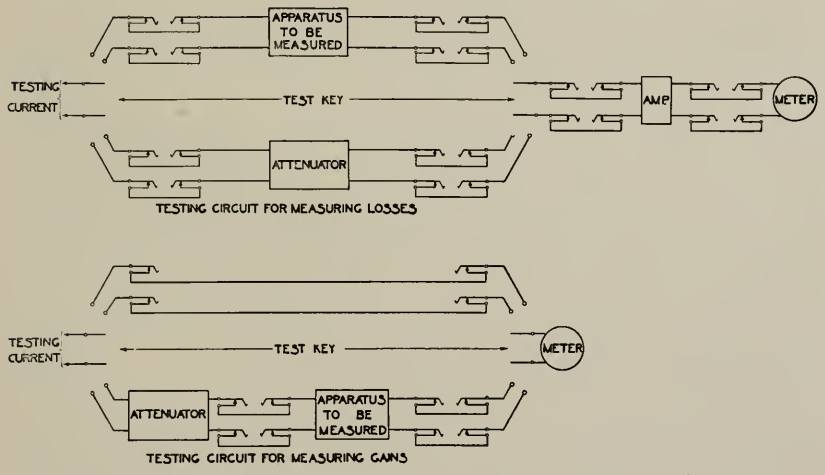


Figure 29—Schematic of testing circuit for type "C" carrier telephone system

carrier-frequency oscillator. Only high-frequency currents are transmitted through the carrier repeater, hence the 1,000-cycle supply employed at the terminals where modulation and demodulation of these carrier currents occur, is not useful in testing the repeater equipment.

It is customary to make periodic tests of the equipment. On the longer circuits [8] each day the channels are "lined up" for the required overall transmission equivalents. At less frequent intervals the vacuum tubes are checked for emission, the gains of the sending and receiving branches are measured, the carrier synchronism is checked, etc.

Vacuum Tubes. Two principal types of vacuum tubes are employed in this system. One of these is the so-called "L" tube. This tube has a filament circuit requiring a current of approximately 0.5 ampere with a voltage drop of 4.0. It has a μ of 6.5, and a normal plate current, when used as an amplifier with a "B" voltage of 130 and a "C" biasing potential of 8, of about 6.5 milliamperes.

For the output stage of the amplifier a higher capacity tube is employed. This is a so-called "O" tube having a μ of about 2.5, a filament current requirement of approximately one ampere at a voltage drop of 4.5. The normal plate current when used as an amplifier with a grid biasing potential of 22 and a plate potential of 130 volts is from 17 to 35 milliamperes.

In addition to the above the pilot channel uses a low-filament current tube. This tube has a μ of 8, a filament current of .060 ampere, and a filament voltage drop of 3 volts. The normal plate current when used as an amplifier with a grid biasing potential of 7 volts and a plate potential of 130 volts is approximately .003 ampere.

The tubes employ oxide-coated filaments and have been designed to be especially long lived to meet daily 24-hour service requirements.

Power Supply. The power required for the carrier equipment is taken from the telephone office supply where this is adequate. Usually the 24-volt central office power plant is suitable for the purpose, and the 130-volt supply for the plate circuits of the tubes is taken from the same batteries provided for telephone repeaters if available. The carrier requirements, however, may amount to a substantial addition to the load on the power plant, particularly where several carrier systems are installed in a relatively small office.

The amount of power required for a typical carrier telephone system is substantially larger than the usual telephone power requirements for the same number of facilities. Each system terminal requires approximately 8 amperes at 24 volts and about 400 milliamperes at 130 volts. The power required for each carrier repeater amounts to about 4 amperes at 24 volts and 250 milliamperes at 130 volts. Thus, the total power required for one three-channel

[8] W. H. Harden, "Practices in Telephone Transmission Maintenance Work," *Bell System Tech. Jl.*, V. 4, Jan. 1925, pp. 26–51.

system with two intermediate repeater stations would be in the neighborhood of 24 amperes at 24 volts and 1.3 amperes at 130 volts, amounting altogether to about 750 watts. This corresponds roughly to the amount of power consumed by about 80 telephone repeaters, so that the total power required for three such carrier systems would be about equal to that required for a cable repeater station having between 200 and 300 repeaters. Assuming 2 or 3 repeaters in a typical voice circuit, the carrier systems are seen to require over ten times as much power as voice circuits in providing the same facilities.

The best results are obtained with the carrier systems when very close regulation of this power supply is maintained. About $\pm$ 1 volt for the 24-volt supply and $\pm$ 5 volts for the 130-volt supply are desirable limits of variation. Means for obtaining such regulation are added as required to the existing power plant. In the larger offices this may consist of a duplicate battery with full-floating operation. In the smaller installations, a relay regulating circuit may be added which controls the filament current as the voltage varies from 20 to 28 volts. This consists of a sensitive voltmeter relay arranged with accessory relays to cut resistance in and out of the individual filament supply circuits as the voltage varies.

Design of Carrier Apparatus

Many will, no doubt, be interested in the further technical details of some of the more important units of the carrier system.

In the development of the apparatus considerable preliminary work was necessary to determine the circuit requirements imposed upon the individual units. For example, preliminary to the design of the filters, laboratory studies were made to find what interfering frequencies might be expected in a channel and what attenuation the different filters must offer at various points in the frequency range, in order that the system should provide speech of satisfactory quality and freedom from interference. As a result of such work, it was possible to make the requirements of the filters no more stringent than absolutely necessary, thus keeping the cost down to a minimum while insuring adequate performance. Preliminary studies were also made on the other parts of the system such as modulator, demodulator, oscillator, etc. In the descriptions which follow, no attempt has been made to describe this preliminary work, the discussion being limited to the requirements imposed, and the circuits devised to meet these requirements.

Modulator and Transmitting Oscillator. A circuit drawing of the modulator is shown in Figure 30. It may be considered that the

function of the modulator is to translate the voice frequencies along a frequency scale to some assigned location in the band of frequencies to be occupied by the carrier system. The carrier frequency controls

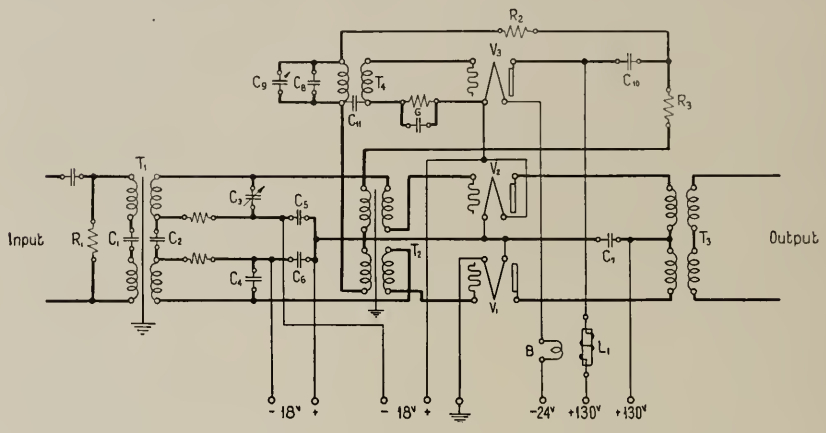


Figure 30—Schematic of modulator circuit and transmitting oscillator

the location of the shifted voice band or sideband, and the oscillator, which provides this frequency, is made an integral part of the modulator unit. The modulating process requires a circuit element having a non-linear response characteristic to produce the sideband from a combination of the carrier and voice frequencies. In this circuit the three-element vacuum tube is operated to give this required characteristic.

Since the carrier is not transmitted in the type "C" system, each modulator and demodulator becomes a complete and independent frequency changing unit. As previously noted, the frequency stability of each oscillator must be sufficiently good so that the carriers of the corresponding modulator and demodulator units will differ but slightly in frequency so as not to affect unfavorably the speech which is transmitted. In this connection both naturalness of the received speech and intelligibility must be considered. In the type "C" system satisfactory results have been obtained by holding the difference between the carrier frequencies of any two associated units due to all causes to within about 20 cycles.

The usual causes of frequency variation are fluctuations in the A and B battery supply and changes in temperature and humidity. Vacuum tubes have to be replaced periodically and differences in the tube characteristics may cause a slight variation in the frequency.

The type of oscillator circuit was chosen to furnish the greatest

frequency stability with variations in power supply, particularly in the plate battery. Fluctuations in the filament current are reduced by the use of ballast resistors to a point where they do not appreciably affect the oscillator frequency.

In order to maintain stability with temperature and humidity changes, it was necessary to develop circuit elements (primarily the inductance in the oscillating circuit) which were not greatly affected by these variables. As a result the oscillators vary less than 10 cycles per second at the highest carrier frequency with power variations within the limits of plant maintenance, and have a frequency temperature coefficient of approximately .002 per cent per degree Fahrenheit. This corresponds to about one cycle per second per degree Fahrenheit in the highest frequency units used in the type "C" system. The temperature difference between offices containing terminal equipment seldom exceeds 20 degrees Fahrenheit.

An extreme change in frequency of 20 cycles per second may be encountered with different tubes. Maintenance experience has shown that it is usually unnecessary to check the synchronization of the modulator and demodulator carrier frequencies more often than once a week, unless tubes are replaced or some other unusual circuit change occurs.

The modulating tubes are placed in a push-pull arrangement, and the carrier voltage is applied to both grids in phase and the suppression of the carrier frequency secured by a differential connection of the output transformer windings. It is difficult to completely suppress the carrier and a limit is set upon the amount which can be allowed on the high-frequency line without causing interference between systems. This limit requires that the carrier flowing out from the modulator should not exceed approximately 500 microamperes. With varying conditions of power, the balance of the modulator cannot be maintained absolutely constant, so that to insure meeting this requirement under the worst conditions it is necessary to adjust the balance under normal conditions to a point where the carrier has been reduced to about 150 microamperes at the output of the modulator. The sideband current flowing at this place in the circuit is ordinarily of the order of 2,000 microamperes. Adjustment of the carrier balance is made by changing the condenser across one half of the input circuit and by selecting tubes.

A further requirement imposed on the modulator unit is that of gain stability. In order to maintain sufficiently constant transmission over a circuit there must be a high degree of inherent stability in all those units whose variations are not included in the indications of

the pilot channel. The modulator and demodulator are the only units in this category.

In the modulator and demodulator the principal factors tending to cause instability of gain are the variations in plate and filament battery where these occur and changes in the tubes during their life. A characteristic behavior of the modulator described below has been used to advantage in minimizing this instability. The gain of the modulator for varying values of the carrier voltage passes through a maximum near the point where the grids of the modulator tubes are driven to a positive potential, with respect to the filament. The output from the carrier oscillator will increase with increasing plate potential, and due to the above characteristic may be made to compensate somewhat for the tendency of the gain of the modulating tubes to increase. Figure 31 shows the change in gain of the modu-

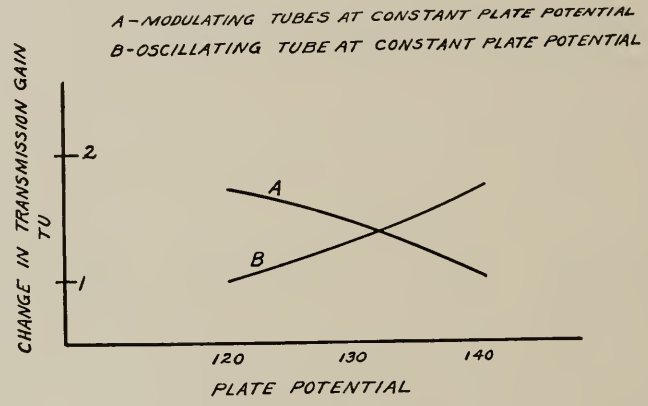


Figure 31—Modulator gain—relation to plate potential

lator circuit when the plate potential of either the oscillating or modulating tubes is changed independently. With these arrangements the total variation in the modulator or demodulator gain, due to the fluctuations of power supply, does not usually exceed ± .25 TU. The possible variation due to tube differences is somewhat larger, approximately ± .7 TU. This is not serious since tubes are ordinarily replaced when a system is out of service, and the gain can be readjusted before the system is restored to operation.

Another requirement which the modulator must meet is one of transmission quality or equality in transmission gain at various frequencies in the voice band. The modulator should not limit the band of frequencies which are to be transmitted over the system. In other words, the gain of the modulator should be substantially the

same for frequencies between 200 and 2,800 cycles per second. The characteristics of the transformers and the impedance in which the output of the modulator is terminated are the controlling factors in the quality of the modulator. A typical characteristic of modulator gain with frequency under ideal terminations is shown in Figure 32.

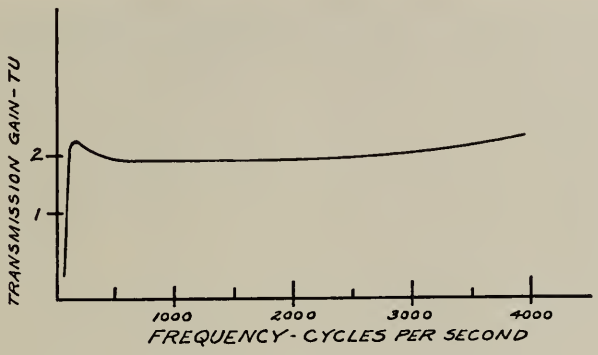


Figure 32—Modulator gain—relation to frequency

In the type "C" system the effect of the band filter impedance is to cause a variation in this characteristic of approximately 5 TU at 2,800 cycles.

The final requirement placed upon the modulator relates to the energy level which must be handled. It should not be possible to overload the modulator seriously with the amount of power produced by a subscriber's set at the transmitting toll testboard level. With a given modulator circuit, this requirement can be met by designing the input transformer with the proper turns ratio.

The following paragraphs give a more detailed description of the actual circuit which has been developed to meet the above requirements.

The voice-frequency circuit is through the hybrid coil to the terminals of the input transformer. The resistance placed across the input circuit is to improve the impedance terminating this branch of the hybrid coil, and thus improve the terminal impedance looking into the hybrid coil from the voice-frequency line. The condenser C-1 is inserted in series with the primary winding of the transformer to improve the transmission characteristic of the circuit at low frequency. This condenser resonates with the inductance of the primary winding, increasing the voltage across the primary at low frequencies where the modulator input circuit tends to become less efficient. The two windings of the secondary side of the transformer are separated

by by-pass condenser C-2, as they are at different potentials from ground, due to the series connection of the filaments of the vacuum tubes. The variable condenser C-3 affords a means of balancing the carrier frequency potentials. Condenser C-4 is one half the maximum capacity of C-3 in order that carrier frequency unbalance



Figure 33—Assembly of modulator panel. (Front view)

in either side of the circuit may be compensated for to a sufficient degree by means of the one adjustable condenser C-3. The voltages, E_1 and E_2, provide the grid bias for the modulating tubes V-1 and V-2. The condensers C-5 and C-6 provide a low impedance path around the source of biasing potentials for the carrier frequency, and condenser C-7 in the plate circuit performs the same function with respect to the plate battery.

In the oscillator, the condensers C-8 and C-9 together with the inductance of one winding of the transformer T-4 form the oscillating circuit. C-9 is made adjustable to compensate for manufacturing variations in the inductance, and to provide in addition a certain flexibility in frequency adjustment. A grid bias for the oscillating tube is provided by the grid leak-condenser combination C. The plate battery is connected through the retardation coil L-1, which presents a high impedance to the carrier frequencies, and prevents

them from flowing through the plate battery. The carrier current in the plate circuit divides between two paths, one through R-2, the feedback resistance to the grid circuit, and the other through R-3, the output resistance, and the transformer T-2 which impresses the carrier voltage on the grids of the modulating tubes. The filaments of the tubes in the modulator circuit are wired in series, and the current flow is regulated by a ballast resister B.

Figures 33 and 34 show the front and rear views of the modulator panel. The adjustable condensers which control the carrier frequency

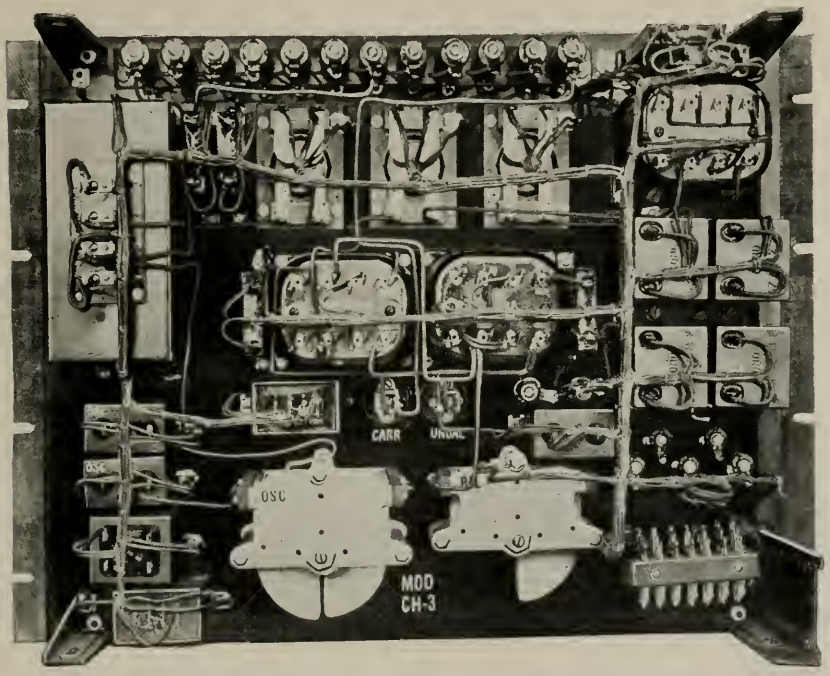


Figure 34—Assembly of modulator panel. (Rear view)

and the carrier balance are accessible from the front of the panel. In the rear view, the oscillator circuit occupies the left-hand side of the picture. The oscillating transformer is in the upper left-hand corner, with the oscillating condensers directly below it. The feedback and output resistances are connected across the top of the panel. The oscillating tube is left of the three tubes, and the carrier input transformer is below it. The voice input transformer is to the right of the carrier transformer, and the output transformer is located in the upper right-hand corner. A metal cover fits over the complete panel at the back to provide electrical shielding and mechanical

protection. All outside connections to the panel are made through the terminal block in the lower right-hand corner. Wires supplying power, together with those which are at a low a.-c. potential with respect to ground, are run in a cable, while wires at a high a.-c. potential are run directly from point to point in as short a path as possible in order to reduce losses resulting from the capacity of these wires to ground.

Demodulator and Receiving Oscillator. The circuit of the demodulator shown in Figure 35 is in many respects similar to that of the

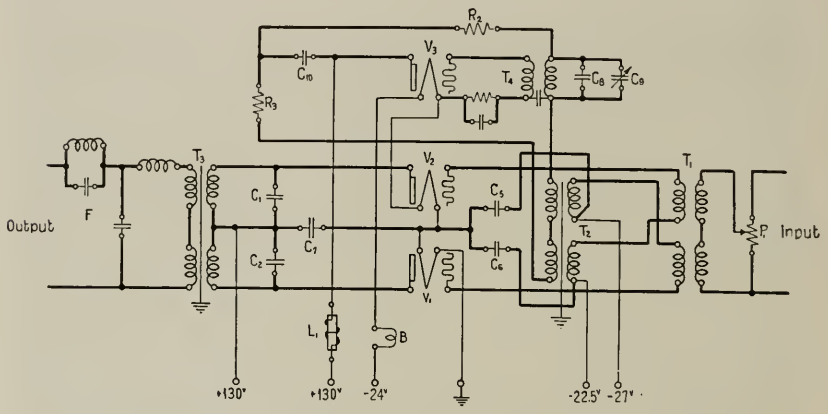


Figure 35—Schematic of demodulator circuit and receiving oscillator

modulator. The function performed by the demodulator is also similar, being a translation from a high-frequency band to a lower instead of the reverse.

The oscillator which supplies the carrier to the demodulator is of the same type as the modulator oscillator, and has been discussed in connection with that circuit. No adjustable feature for balancing the carrier is required in the demodulator circuit. The carrier suppression needed in addition to the suppression inherent in the balanced circuit is provided by the low-pass filter at the output. If the carrier is not sufficiently suppressed, it will pass into the voice circuit or across the hybrid coil into the associated modulator, causing in some channels an objectionable beat tone.

The transmission stability of the demodulator is obtained by the same methods used in the modulator since the performance of the two circuits is similar, and the transmission quality requirement is essentially the same for both units. A typical demodulator characteristic is shown in Figure 36. This characteristic at the higher frequencies is controlled by the low-pass filter.

One feature which is required with the demodulator, but not with the modulator, is a variable control of the transmission gain of the circuit. Due to the unequalized transmission of the line section

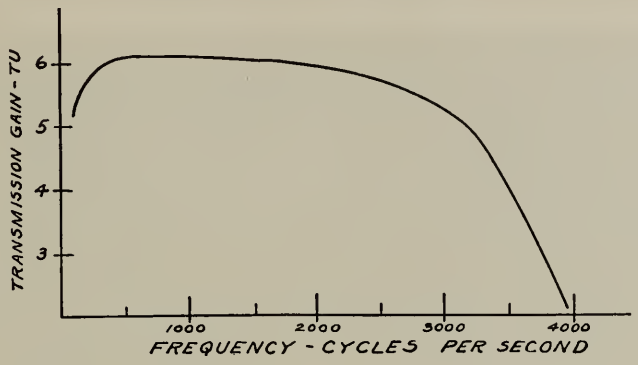


Figure 36—Demodulator characteristic—gain and frequency

adjacent to the terminal, or other differences in the channel equivalents, the three sideband currents normally arrive at a receiving terminal with unequal strength. A potentiometer controlling the gain of the demodulator permits of an equalization of the overall losses on the three channels.

In the following detailed description of the demodulator circuit other minor differences between it and the modulator may be pointed out:

The sideband frequencies enter the demodulator passing to the potentiometer P-1 which controls the amount of current to the input transformer T-1. The position of the carrier input transformer T-2 is somewhat different in the demodulator circuit as compared to the modulator circuit, due to the difference in the high-frequency characteristic of the T-1 transformers. In the modulator this transformer must be designed to transmit voice frequencies primarily. It has a comparatively large capacity to ground which would reduce the effective carrier voltage on the tube grids if it were placed in the same circuit position as is the demodulator transformer. The function of most of the circuit elements is evident from the previous description of the modulator. The C-1 and C-2 condensers provide a low impedance path for the carrier frequency. They are necessary here because the transformer T-3 designed for high efficiency at voice frequencies has considerable leakage inductance, which would present a high impedance to the carrier in the plate circuit if the condensers were not provided. For the maximum gain the impedance of this

circuit should, of course, be a minimum at carrier and sideband frequencies. At the output a low-pass filter structure F provides for the suppression of the unwanted products of demodulation.

A front view of the demodulator unit is shown in Figure 37. The



Figure 37—Assembly of demodulator panel. (Front view)

panel layout and general appearance is similar to that of the modulator. The two dials shown control the demodulator input and the oscillator frequency, respectively, as indicated in these figures.

Filters. The general function of a band filter is the selection of a band of frequencies, and the protection of this band from interfering frequencies located on either side. The filters determine what band width is transmitted, and thus to that extent they control the quality of speech which may be obtained through the carrier circuit. The type "C" system transmits a band corresponding to approximately 200 to 2,700 cycles per second in the voice range.

In considering the requirements imposed upon the band filters it is necessary to keep in mind [9] the fact that the modulator produces not only the particular sideband which is to be transmitted but also an unwanted sideband of the same volume as the wanted sideband and

[9] R. V. L. Hartley, "Relation of Carrier and Side Bands in Radio Transmission," *Bell System Tech. Jl.*, V. 2, April 1923, pp. 90–112.

equal to it in width, located on the opposite side of the carrier frequency. In addition to these products of modulation there are produced other frequency bands, the important ones occupying sideband positions about the harmonics of the carrier frequency. See Figure 38.

The first requirement on the band filters is imposed by the need of suppressing the unwanted sideband to prevent distortion when the

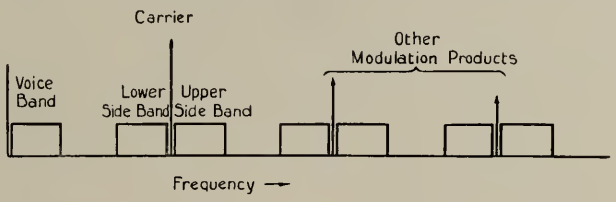


Figure 38—Frequency range of products of modulation

carriers are out of synchronism. The tests mentioned above in connection with the oscillator frequency stability were made with but one sideband transmitted. If both sidebands are transmitted, the carriers must be exactly in synchronism or a "wobble" due to the demodulation of both sidebands can be detected. One sideband must be suppressed by an increasing amount as this carrier difference increases. For a carrier frequency difference of about 20 cycles it is necessary to suppress the unwanted sideband about 40 TU, thus reducing it to about 1/100 of the strength of the wanted sideband in order to eliminate completely this type of distortion. This requirement can be met by providing the necessary attenuation in either the transmitting or the receiving band filter, or by making the sum of their attenuations equal to 40 TU.

The suppression of the unwanted sideband is necessary for another reason in a multi-channel system in which the transmitted sidebands are close together. The unwanted sideband from one channel overlaps the wanted sideband of an adjacent channel, and would be demodulated and appear as "crosstalk" into this channel if it were not suppressed by the transmitting band filter. The suppression needed is determined by the amount of interference which can be tolerated from one channel to another. It has been found that to meet this requirement the transmitting band filter must suppress the unwanted sideband about 60 TU. The other modulation products mentioned above must also be reduced by the transmitting band filter to a value which will not cause interference in any channel into which they might pass. The discrimination requirement for these frequencies is less severe because the magnitude of these modulator products is not so great.

A particular termination is required at the end of the filter which is connected to the modulator. In order to get the maximum sideband power out of the modulator used, the impedance of the associated band filter, seen from the modulator, must be made low over the range of voice frequencies.

With the channels placed closely together and with the coordination of different types of systems, depending upon the channel locations, it is important that the band filters remain constant after manufacturing, and that all filters of the same type be manufactured to meet close requirements. For proper coordination between systems it has been found desirable to keep all the channel bands within $\pm$ 125 cycles of an assigned location. This means in the higher frequency channels that the filters must be manufactured to a frequency accuracy of the order of 1/2 of 1 per cent.

The attenuation requirements for the receiving band filter are somewhat different from those of the transmitting band filter. The purpose of the receiving band filter is the suppression of the frequencies of the adjacent channels as they are received over the line. In contrast to the transmitting filter, which must suppress the unwanted frequencies produced in its own channel, a filter with somewhat different characteristics could, therefore, be used for a receiving filter. While the requirements were determined separately for the receiving and transmitting filters, it was desirable in the interest of manufacturing economy to build both alike, setting requirements on the basis of a double purpose filter. Thus, this filter had to provide attenuation at each frequency to meet the more severe of the requirements for either the transmitting or the receiving position. Figure 39 shows the transmitting characteristic of a typical filter designed to meet the requirements outlined above.

As has been explained, the grouping of the channel bands in opposite directions requires the use of so-called directional filters at terminal and repeater points. These filters occur in the circuit in pairs—each pair consisting of one high-pass and one low-pass filter. The "cut-off" point of the filters is determined by the type of system in use—C-S or C-N and its corresponding "grouping point." At repeater points the filters are split for each direction in order to provide selectivity at both the output and input circuits of the amplifiers.

Considering the closed circuit through the two amplifiers and the four directional filters, the attenuation in this loop must be considerably greater than the sum of the gains of the two amplifiers at all frequencies. In the regions outside of the carrier frequencies, the margin between attenuation and gain is made about 10 T. U. For

frequencies in the carrier range this margin must be still greater to prevent distortion, which becomes objectionable when circulating currents of any size are allowed to exist. This "feed-back" effect

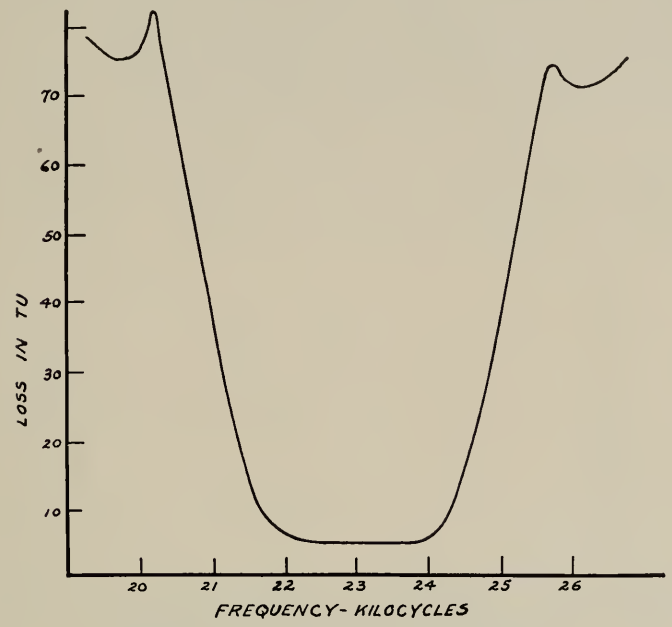


Figure 39—Typical band filter characteristic

will also affect the repeater input impedance, and because of the necessity for closely controlling this characteristic the margin between gain and attenuation is not permitted to be less than 25 T. U. at any frequency used for transmission in either direction. The impedance of these filters on the line side must match the line impedance closely in order that no considerable reflection of the carrier currents can take place at this junction point.

As was mentioned previously, the output of an amplifier contains, due to modulation, other frequencies in addition to those which compose the input, so that crosstalk is to be expected between some of the channels. The amount of this crosstalk, which will appear at the far end, depends on the ratio of the sideband currents to the interfering currents produced in the amplifier, the measurement being made at the repeater output. The near-end crosstalk, however, is dependent on the level difference between the strong output of the one amplifier and the weak input to the other. Those frequencies which may give trouble in the channels at the near end enter the

returning circuit at the amplifier input, a point where the sideband level is very low. To put the near-end crosstalk on the same basis as the far end, the output directional filter must introduce enough attenuation in its non-transmitting range to make up this level difference. This attenuation is increased until the near-end crosstalk due to this cause is appreciably less than the far end.

The output current of one amplifier may be 30 TU or more stronger than the input current to the amplifier for the opposite direction, and the directional filter at the input of this second amplifier must offer sufficient attenuation to the output currents of the first so that they will not contribute materially to its load.

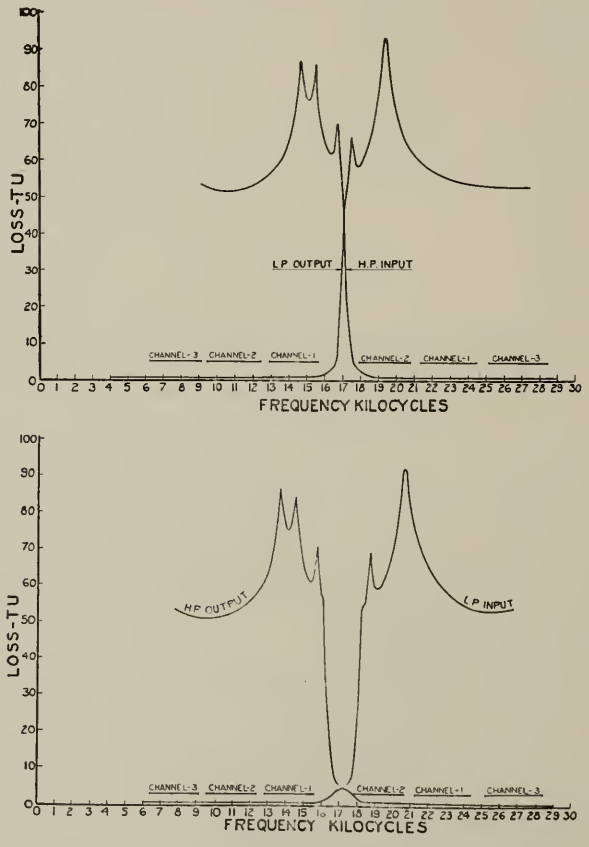


Figure 40—Typical directional filter characteristics

Figure 40 shows the selectivity characteristics of the two directional filters.

A pair of filters having important functions is the line filter set

which, as has been noted, acts to separate the carrier currents from the regular speech currents on the common line circuit. It consists of a high-pass and a low-pass filter paralleled on the line side. Currents entering these terminals from the line circuit pass through the high-pass circuit to the carrier apparatus or through the low-pass circuit to the circuit terminal or repeater. The transmission characteristics of these filters are shown on Figure 41. It will be noted that frequencies

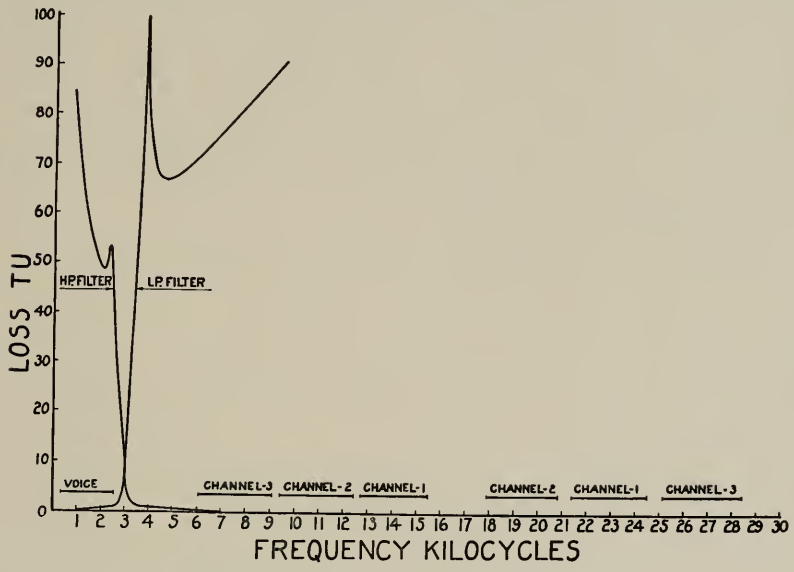


Figure 41—Typical line filter characteristics

above approximately 3,300 cycles are transmitted in the high-pass circuit and frequencies below about 2,800 cycles are transmitted through the low-pass circuit. It is common to equip a few line circuits with line filter sets, in addition to those which are normally in use for carrier transmission. This makes it readily possible in case of an emergency or for other reasons to use the spare wires thus equipped for carrier transmission.

Non-linear effects may be produced in the coils and condensers in the circuit. The design of the filter parts must be made so that these effects will be a minimum. This requires the use of non-magnetic cores in the coils, and also that the containers be of non-magnetic material. Condensers in magnetic containers must be located so that they will not lie in the field of the coils and thus contribute to the modulation products. The modulation in the line filters, telegraph composite sets, and office and cable loading units, must also be considered.

As already mentioned, care has to be exercised in the mounting of filters belonging to different systems in the same office, so that no crosstalk will be introduced from one system into another. A considerable level difference may exist between two filters of different systems, and it may be desired to mount these filters on adjacent bays. In order that the crosstalk between these two systems may be kept within desirable limits, the separation between the filters must, in some instances, correspond in attenuation loss to the order of 120 TU, or one part in a million. To meet this exacting requirement, the filters are totally incased in sealed copper boxes, the leads being brought out through small holes to terminal blocks.

Amplifiers. As previously mentioned, the amplifiers employed with the type "C" system at the terminals are identical with those used with the repeaters at intermediate stations. The following is, therefore, applicable to both cases:

The number and size of tubes needed to deliver the necessary output level or power are largely controlled by interchannel crosstalk requirements. With the grouping frequency arrangement, the three bands which transmit in the same direction are amplified in a common circuit. The different sideband frequencies in passing through the common amplifier must not react upon each other to produce other frequencies of sufficient magnitude to cause interference. For example, second harmonics of the lowest band frequencies lie within the range of the highest channel in the lower group. If these harmonics are permitted to become too great, troublesome noise will be present in the highest channel when speech currents flow in the lowest. In order that this interference or crosstalk may not become excessive the tubes used in this amplifier must be made of ample power capacity.

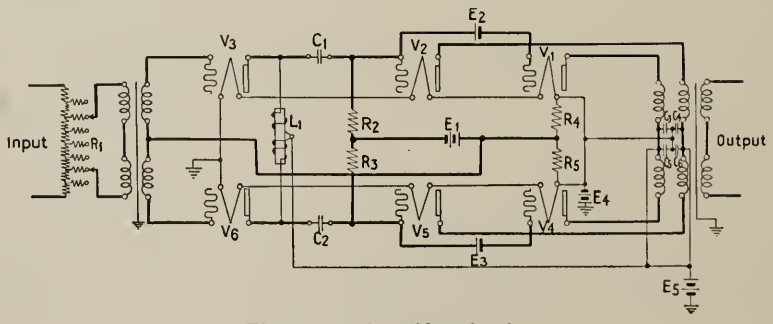


Figure 42—Amplifier circuit

This example of interference caused by the second harmonic shows the desirability of using a push-pull amplifier in carrier repeaters

because of its property of balancing out second order effects, which in a single tube or unbalanced circuit are the largest of all the modulation products at the usual loads.

The currents from the three channels enter the carrier amplifier shown in Figure 42. The circuit consists of two stages; the first stage of two tubes, the second of four of higher power rating. The gain is controlled in 2 TU steps by the adjustable potentiometer in the input. The gain frequency characteristics for different potentiometer settings are substantially flat within a small fraction of a TU over the range of any channel.

The amplifiers for the two directions are of slightly different design, each amplifier being arranged for a flat characteristic over its own group of frequencies. It has been stated that the load capacity of the amplifier is limited [10] by the modulation products which increase with the load. Figure 43 shows the amount of second and third

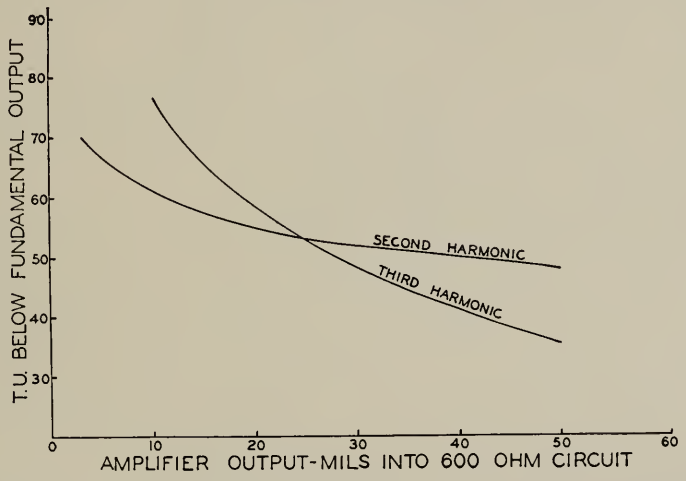


Figure 43—Amount of second and third harmonics as function of carrier repeater output

harmonics produced in a typical repeater with varying single frequency output. By connecting the tubes in push-pull instead of in parallel, the second harmonics have been reduced by about 15–20 TU. Other products of modulation as well as the second and third harmonics increase with the output and thus the power which can be taken from the amplifier under the operating conditions is limited as these effects are likely to result in interchannel interference.

When the alternating voltage applied to one grid is positive with

[10] F. C. Willis and L. E. Melhuish, "Load Carrying Capacity of Amplifiers," *Bell System Tech. Jl.*, V. 5, October 1926, pp. 573–592.

respect to the filament, that on the other grid is negative. Since the even order products are proportional to an even power of the input voltage, these currents will flow through the high side winding of the output transformer non-inductively producing no flux in the transformer, and hence no current in the low side windings. To realize this ideal condition, the two currents flowing in the output transformer windings must be equal in amplitude, and 180 degrees out of phase. Like amplitudes can be obtained in several ways since the plate current is a function of a number of tube constants. Tubes may, therefore, be selected which will give the same harmonic current, that is, tubes in which the net effect of the several factors is the same.

Conclusion

Use in Telephone Plant. The carrier systems are meeting successfully and economically the requirements of long distance telephone service. From what has already been written, it is evident, however, that the apparatus is by its nature complex and to a fair degree expensive, so that for the relatively short distances it is cheaper to string additional wire. The exact distance beyond which it is more economical to employ carrier methods is obviously dependent on the circumstances surrounding each particular case. Systems are operating for distances of 150 miles and upwards.

Traffic growth often requires additional circuits for the shorter distances, where there are longer haul continuous physical circuits on the same line. In this case it is not uncommon to break up the long haul physical circuits into sections to satisfy the short haul circuit growth and to install a carrier system to meet the long haul needs.

The growth of the use of carrier systems has already been pictured. How the systems are distributed over the lines of the Bell System is shown on Figure 44. The heaviest density of use occurs in the middle and western sections and in general where the circuit demand and growth have not reached the large figures required to justify the installation of toll cables. In particular, the section west of the Mississippi is a promising field for the application of carrier systems.

Future. While the type "C" system satisfies those circuit growth demands for moderate and long haul, there has remained undeveloped a considerable field for carrier methods over the shorter distances where only wire stringing has hitherto been economical. Very recent developments have resulted in the trial and early field applications of a simple single-channel carrier telephone system designed particularly to meet these shorter haul demands and thereby to secure the greatest practicable economy in providing facilities by carrier methods in the

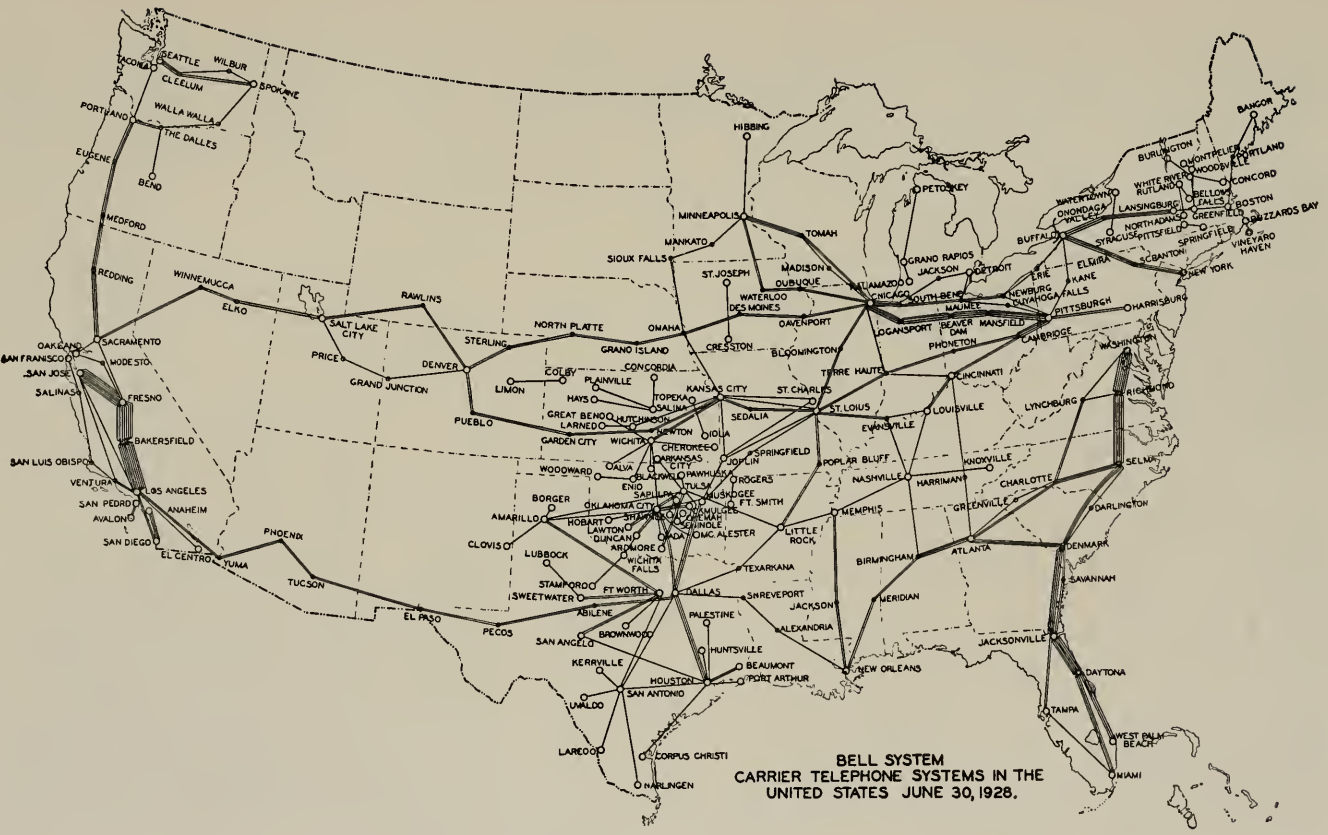


Figure 44—Map showing carrier telephone systems throughout Bell System

Bell System. It is naturally finding its most extensive use in the sections of the telephone plant where the shorter circuits predominate. Because of the fact that the type "D" system development is only now being completed it was thought desirable in the present paper to confine attention to the long haul system (type "C") and to defer the presentation of the detailed information on the short haul development until a somewhat later date.

While considerable progress has been made in the development and application of these carrier systems since the beginning of their use about ten years ago, there is still much to be done in the matter of simplifications and further use of the high-frequency spectrum. Automatic pilot channel arrangements are being tested whereby manual maintenance costs can be reduced. Further developments are anticipated in the matter of transposition arrangements to permit an open-wire line to carry multi-channel long haul carrier systems on most of its pairs. While the systems now in use in the field employ frequencies no higher than approximately 30,000 cycles, frequencies considerably higher than this can undoubtedly be economically employed.

Bibliography

The following more important papers and articles relating to multiplex carrier telephony and telegraphy are offered, with no pretensions as to completeness, as an addition to the bibliography which forms part of the Colpitts-Blackwell paper. Reference should, therefore, be made to the latter for articles published prior to 1921.

H. S. Osborne. The Design of Transpositions for Parallel Power and Telephone Circuits. *A. I. E. E. Transactions*, V. 37, June 1918, pp. 897–936.

E. H. Colpitts and O. B. Blackwell. Carrier Current Telephony and Telegraphy. *A. I. E. E. Transactions*, V. 40, 1921, pp. 205–300.

G. A. Culver. Guided Wave Telephony. *Jl. of Franklin Institute*, V. 191, March 1921, pp. 301–328.

R. A. Heising. Modulation in Radio Telephony. *I. R. E. Proceedings*, V. 9, August 1921, pp. 305–352.

G. A. Campbell. Physical Theory of the Electric Wave-Filter. *Bell System Tech. Jl.*, V. 1, No. 2, November 1922, pp. 1–32.

O. J. Zobel. Theory and Design of Uniform and Composite Electric Wave Filters. *Bell System Tech. Jl.*, V. 2, No. 1, January 1923, pp. 1–46.

R. V. L. Hartley. Relation of Carrier and Side Bands in Radio Transmission. *Bell System Tech. Jl.*, V. 2, April 1923, pp. 90–112, and *Proceedings of I. R. E.*, V. 11, No. 1, February 1923, pp. 34–56.

A. F. Rose. Practical Application of Carrier Telephone and Telegraph in the Bell System. *Bell System Tech. Jl.*, V. 2, No. 2, April 1923, pp. 41–52.

Paolo Borgatti. Telefonia e Telegrafia ad Alta-Frequenza au Fili. *L'Elettrotecnica*, V. 10, August 16, 1923, pp. 531–541.

H. S. Osborne. Telephone Transmission over Long Distances. *A. I. E. E. Jl.*, V. 42, No. 10, October 1923, pp. 1051–1062.

Rudolph Fiedler. Die Ruhmerschen Hochfrequenz Apparate fur Mehrfachferusprechen auf Leitungen im Reichspostmuseum in Berlin. *Archiv f. Post u. Teleg.*, V. 52, 1924, pp. 123.

Rudolph Fiedler. Die Aussichten der Leitungsgerichteten Hochfrequenztelephonie u. -Telegraphie in Deutschland. *Teleg. u. Fernsprech*, V. 1, January 1924, pp. 5–8.

Hans Muth. Mehrfach Telephonie und Telegraphie Langs Leitungen. *Telef. Zeit.*, V. 6, No. 34, January 1924, pp. 27–44.

N. H. Slaughter and W. V. Wolfe. Carrier Telephony on Power Lines. *A. I. E. E. Jl.*, V. 43, No. 4, April 1924, pp. 377–381.

H. Schulz and K. W. Wagner. Der Drahtfunk. *E. T. Z.*, V. 45, May 15, 1924, pp. 485–489.

R. V. L. Hartley. The Transmission Unit. *Electrical Communication*, V. 3, No. 1, July 1924, pp. 34–42.

F. B. Jewett. Making the Most of the Line. *Electrical Communication*, V. 3, July 1924, pp. 8–21.

W. H. MARTIN. Transmission Unit and Telephone Transmission Reference Systems. *Bell System Tech. Jl.*, V. 3, July 1924, pp. 400–408, and *A. I. E. E. Jl.*, V. 43, No. 6, June 1924, pp. 504–507.
LOUIS COHEN. Radio Sur Lignes. *L'Onde Elect.*, V. 3, No. 34, October 1924, pp. 477–490.
N. H. SLAUGHTER. Carrier Telephone Equipment for High-Voltage Power Transmission Lines. *Electrical Communication*, V. 3, No. 2, October 1924, pp. 127–133.
W. H. HARDEN. Practices in Telephone Transmission Maintenance Work. *Bell System Tech. Jl.*, V. 4, No. 1, January 1925, pp. 26–51.
W. V. WOLFE. Carrier Telephony on High-Voltage Power Lines. *Bell System Tech. Jl.*, V. 4, No. 1, January 1925, pp. 152–177.
B. P. HAMILTON, H. NYQUIST, M. B. LONG AND W. A. PHELPS. Voice Frequency Carrier Telegraph System for Cables. *A. I. E. E. Jl.*, V. 44, No. 3, March 1925, pp. 327–332, and *Electrical Communication*, V. 3, No. 4, April 1925, pp. 288–294.
J. J. JAKOSKY AND D. H. ZELLERS. Line Radio and the Effects of Metallic Conductors on Underground Communication. Reports of Investigations, Department of the Interior—Bureau of Mines, Serial No. 2682, April 1925.
R. A. HEISING. Production of Single Sideband for Transatlantic Radio Telephony. *I. R. E. Proceedings*, V. 13, No. 3, June 1925, pp. 291–312.
W. H. CAPEN. Wire Telephone Communication in Theory and Practice. *Electrical Communication*, V. 4, January 1926, pp. 161–183.
THOMAS SHAW AND WILLIAM FONDILLER. Development and Application of Loading for Telephone Circuits. *Bell System Tech. Jl.*, April 1926, pp. 221–281, and *A. I. E. E. Jl.*, V. 45, No. 3, March 1926, pp. 253–263.
B. R. CUMMINGS. Carrier Current Communication. *General Electric Review*, V. 24, May 1926, pp. 365–367.
H. W. HITCHCOCK. Carrier Current Communication on Submarine Cables. *A. I. E. E. Jl.*, V. 46, No. 10, October 1926, pp. 923–929, and *Bell System Tech. Jl.*, V. 5, October 1926, pp. 636–651.
F. C. WILLIS AND L. E. MELHUISH. Load Carrying Capacity of Amplifiers. *Bell System Tech. Jl.*, V. 5, October 1926, pp. 573–592.
H. A. AFFEL AND J. T. O'LEARY. High-Frequency Measurements of Communication Lines. *A. I. E. E. Transactions*, V. 44, 1927, pp. 504–513.
E. I. GREEN. Methods of Measuring the Insulation of Telephone Lines at High Frequencies. *A. I. E. E. Transactions*, V. 44, 1927, pp. 514–518.
W. J. SHACKELTON AND J. G. FERGUSON. High Frequency Measurements of Communication Apparatus. *A. I. E. E. Transactions*, V. 44, 1927, pp. 519.
W. J. SHACKELTON. Shielding Bridge for Inductive Impedance Measurements at Speech and Carrier Frequencies. *A. I. E. E. Jl.*, V. 44, No. 2, February 1927, pp. 159–166.
G. E. STEWART AND C. F. WHITNEY. Carrier Current Selector Supervisory Equipment. *A. I. E. E. Jl.*, V. 44, No. 6, June 1927, pp. 588–592.
G. A. BODDIE. Telephone Communication over Power Lines by High-Frequency Currents. *Proceedings of I. R. E.*, V. 15, July 1927, pp. 559–640.
G. A. BODDIE. Use of High-Frequency Currents for Control. *A. I. E. E. Jl.*, V. 44, No. 8, August 1927, pp. 763–769.

Abstracts of Bell System Technical Papers Not Appearing in this Journal

Effect of Grounding on Telephone Interference.[1] J. J. Pilliod. This paper, presented before the Pittsburgh Section of the Association of Iron and Steel Electrical Engineers in February 1928, is a rather complete although non-mathematical presentation of the inductive effects of power lines on nearby communication circuits. The production of noise on the latter circuits and the production of voltages sufficiently high to be prejudicial to the operators and users of these circuits are separately discussed. Comparisons are drawn between the inductive action of grounded and ungrounded power lines. Although free from mathematics, the paper gives a very good outline of the interference problem and points out the many opportunities presented for cooperative effort both in connection with original design and with reduction of interference on existing lines.

A Modification of the Rayleigh Disk Method for Measuring Sound Intensities.[2] L. J. Sivian. The usual procedure is to measure the deflexion of the disk under the influence of a steady sound-field. This paper outlines a procedure which has been found useful when the sound amplitude can be made a suitable function of time. The scheme depends on the fact that the torque which the sound-wave exerts on the disk is a non-linear function of the sound amplitude, being proportional to the square of the air particle velocity. The amplitude of the sound-wave to be measured is modulated with a frequency equal to that of the free vibration of the suspended disk. The measurement requires reading the amplitude of oscillations corresponding to the modulating frequency, rather than a steady deflexion of the disk. The disturbances caused by spurious air currents are largely reduced. In addition, in many practical cases at least, there is a gain in absolute sensitivity. Both theory and experimental verification are given.

Reflection of Electrons by a Crystal of Nickel.[3] C. J. Davisson and L. H. Germer. This is a report of some preliminary results obtained in a new series of experiments in which a beam of electrons exhibits the properties of a beam of waves. In previous experiments (*Phys. Rev.*, 30, 705, 1927) a beam of electrons was directed at normal inci-

[1] *Iron and Steel Engineer*, Vol. V, pages 147–155, April 1928.
[2] *The London, Edinburgh, and Dublin Philosophical Magazine, and Journal of Science*, Vol. 5, No. 29, March 1928, pp. 615–620.
[3] *Proceedings of the National Academy of Sciences*, April 15, 1928, Vol. 14, No. 4, pp. 317–322.

dence against a face of a nickel crystal, and observations were made upon the diffraction beams which issued from the incidence side of the crystal at various critical speeds of bombardment. It was anticipated that if the angle of incidence were made other than zero a beam of electrons would be found issuing from the crystal at a series of critical speeds in the direction of regular reflection, and that the series of critical speeds would change with the incidence angle. This regularly and selectively reflected electron beam which is the analogue of the Bragg x-ray reflection beam has been found, and measurements have been made upon it. In the x-ray phenomenon the wave-length of the reflected beam at maximum intensity is related through a simple formula to the angle of incidence and a dimension of the reflecting crystal. This formula (Bragg's formula) does not obtain in the case of electron reflection because of the refraction of the electrons by the crystal. The departures from the Bragg relation are used to calculate indices of refraction of nickel for electrons of various speeds or wave-lengths.

Introduction to Mathematics of Statistics.[1] R. W. BURGESS. This book (282 pp.) discusses the best elementary methods of statistical analysis from the standpoint of a beginner who has had one year of college mathematics, or some practical statistical experience and the ordinary high school mathematics. "Statistical Analysis" is regarded in this book as the logical process by which large masses of quantitative facts may be classified, summarized, analyzed, and compared so as to yield reliable conclusions.

The topics treated include classification, formation of statistical series, use of ratios and percentages in statistical analysis, meaning and graphic discussion of frequency distributions, averages, index numbers, measures of dispersion, trend lines, analysis of seasonal variation, two-, three-, and four-variable correlation, and the elements of sampling and probability. Emphasis is placed on the type of statistical problems most common in the social sciences, in which the data are subject to a higher degree of variability than in the usual problems in physics or astronomy which require the use of the theory of least squares or the Gaussian "curve of error."

The Use of a Moving Beam of Light to Scan a Scene for Television.[2] F. GRAY. The paper is a discussion of a method of scanning employed in the television system demonstrated a year ago at the Bell Telephone Laboratories. A three-dimensional subject is scanned directly by a

[1] Houghton-Mifflin Company.
[2] *Journal of the Optical Society of America*, Vol. 16, pp. 177–190, March 1928.

moving beam of light to produce a picture current in photoelectric cells. This method permits the use of a very intense transient illumination and more than one large-aperture photoelectric cell to collect reflected light. These two factors give a highly efficient optical system for producing a picture current at a transmitting station. The image seen at a distant station is the same as if light came out of the photoelectric cells to illuminate the subject and a small aperture lens formed an image of the subject for transmission. The television system transmits only the spacial variations of brightness and not the absolute brightness of the view; consequently, an additional steady illumination of a subject does not affect the reproduced image.

Maintaining High Standards in Products.[1] E. D. HALL. This article presents briefly but clearly the essential features of a method of keeping before the management an accurate picture of the relative quality of manufactured products. Defects are grouped into four classes and given demerit grades that represent the seriousness of the fault. Defects found each month are added, reduced to an average value, and plotted on charts which as a reference base use the average quality of the preceding five years. A method of computing averages for an entire line of products is also given.

Probability and Its Engineering Uses.[2] THORNTON C. FRY. This book of 470 pages on the Theory of Probability is written from the standpoint of the engineer. Its earlier chapters deal with the fundamental mathematical concepts that underlie the theory, and its later chapters develop these concepts in the directions of their application to traffic and trunking problems, curve fitting, and atomic physics.

Among the subjects which receive especial emphasis are: the logical standing of attempts to determine the probability of an event by trial; the physical significance of the fundamental distribution laws, such as the Normal, the Binomial, and the Poisson Law; Pearson's criterion for "goodness of fit"; and trunking problems.

Differential Intensity Sensitivity of the Ear for Pure Tones.[3] R. R. RIESZ. The ratio of the minimum perceptible increment in sound intensity to the total intensity, $\Delta E/E$, which is called the differential sensitivity of the ear, was measured as a function of frequency and intensity. Measurements were made over practically the entire range of frequencies and intensities for which the ear is capable of sensation. The method used was that of beating tones, this method giving the

[1] "Manufacturing Industries," Vol. 16, No. 1, pp. 17–19, May 1928.
[2] D. Van Nostrand Company.
[3] *The Physical Review*, May 1928, Vol. 31, No. 5, pp. 867–875.

simplest transition from one intensity to another. The source of sound was a special moving coil telephone receiver having very little distortion, actuated by alternating currents from vacuum tube oscillators. Observations were made on twelve male observers. Average curves show that at any frequency $\Delta E/E$ is practically constant for intensities greater than 10^6 times the threshold intensity; near the auditory threshold $\Delta E/E$ increases. Weber's law holds above this intensity, the value of $\Delta E/E$ = constant lying between 0.05 and 0.15, depending on the frequency. As a function of frequency $\Delta E/E$ is a minimum at about 2500 c.p.s., the minimum being more sharply defined at low sound intensities than it is at high. This frequency corresponds to the region of greatest absolute sensitivity of the ear. Analytical expressions are given (Eqs. (2), (3), (4), and (5)) which represent $\Delta E/E$, within the error of observation, as a function of frequency and intensity. Using these equations, it is calculated that at about 1300 c.p.s. the ear can distinguish 370 separate tones between the threshold of audition and the threshold of feeling.

Use of the Noble Metals for Electrical Contacts.[1] E. F. KINGSBURY. The paper describes the results of an investigation of the behavior of gold, silver and the platinum metals as electrical contacts in communication circuits. Platinum has heretofore been considered the standard although some alloys of the platinum metals have been used in especially severe conditions. The economic situation has, however, encouraged the use of cheaper substitutes. Heretofore, accurate knowledge has not been available concerning the intrinsic merit of other materials. This problem is complicated by the various forms of discharges and mechanical conditions encountered in practice. The resistance, erosion, and transfer of contacts are discussed for a variety of materials under various circuit conditions and in different atmospheres.

Economic Aspects of Engineering Applications of Statistical Methods.[2] W. A. SHEWHART. This note calls attention to possible applications of modern mathematical statistical theory, to research, design, production, inspection, supply, and other engineering problems. Attention is given to certain general types of problems in the solution of which statistical applications have been made, and to the nature of the possible economies effected thereby. It is reasonable to believe that very definite economic advantages can be obtained in any large industry through such applications.

[1] Technical Publication No. 95, A. I. M. M. E., March 1928.
[2] *Journal of the Franklin Institute*, Vol. 205, March 1928, pp. 395–405.

Evaluating Quality in Heat-Treated High-Speed Steel by Means of the Milling Cutter.[1] J. B. Mudge and F. E. Cooney. A test of heat-treated high-speed steel in the form of milling cutters, the variables having been reduced to a minimum, and the dulling point of the cutting edges of the tools determined by a recording wattmeter connected in the circuit of the motor of the milling machine. A "deadline" test resulted instead of the usual "breakdown" test.

It was found that:

Cutters of the same steel hardened by the same method check within limits that are sufficiently close for test purposes.

No cast cutter has been found to give results comparable to standard high-speed steel refined by suitable working. Cutters hardened by patented or salt bath processes have not given results comparable to standard high-speed steel hardened by the open fire method.

A Bridge Method for the Measurement of Inter-Electrode Admittance in Vacuum Tubes.[2] E. T. Hoch. A description is given of the Colpitts-Campbell bridge as applied specifically to the measurement of direct admittances in vacuum tubes. Data are given on several tubes.

On Electrical Fields near Metallic Surfaces.[3] Joseph A. Becker and Donald W. Mueller. When an electron escapes from a metallic surface it passes through fields which tend to pull it back. Applied fields when properly directed partially neutralize the surface fields and hence reduce the work the electron has to do against these fields. That is why i, the thermionic current, increases steadily with F_a, the applied field. Quantitatively $d(\log_{10} i)/dF_a = (11600/2.3T)Xs$, where T is the temperature of the surface and s is the distance from the surface at which the surface field F_s is equal to F_a. Hence the slope of an experimental $\log i$ *vs* F_a curve at any F_a yields the value of s corresponding to F_s. For clean or atomically homogeneous surfaces experiment shows that the only force opposing the escaping electron is due to its image field; for composite surfaces other fields, which are ascribed to the adsorbed ions, are superposed on the image field. For 70 per cent thoriated tungsten this "adsorption field" is very large close to the surface and in a direction to help electrons escape; it decreases rapidly in strength as s increases until it is zero at about 15 atom diameters; here it reverses its direction and then increases in strength till it attains a maximum value of 8000 volts/cm. at 75 atom

[1] *Transactions of the American Society for Steel Treating*, February 1928, Vol. 13, No. 2, pp. 221–239.

[2] *Proceedings of the I. R. E.*, April 1928, Vol. 16, No. 4, pp. 487–493.

[3] *Physical Review*, Vol. 31, No. 3, March 1928, pp. 431–440.

diameters; beyond this distance it decreases steadily. The intense field close to the surface accounts for the decreased work function while the reverse field farther out accounts for the poor saturation at ordinary applied potentials.

The photo-electric long wave-length limit should be shifted toward the red by applied fields. This shift should be particularly noticeable for composite surfaces.

Direct Determination of Rubber in Soft Vulcanized Rubber.[1] A. R. KEMP, W. S. BISHOP, and T. J. LACKNER. A modification of the Wijs method is shown to be suitable for determining the rubber content of vulcanized rubber. A procedure for the direct determination of sulfur combined with rubber is also outlined, and the effect of compounding ingredients is shown.

Results of analyses of four reclaimed rubbers by the proposed and difference methods are given for comparison.

Photomicrography and Its Application to Mechanical Engineering.[2] FRANCIS F. LUCAS. This paper was presented at the annual meeting of the American Society of Mechanical Engineers during the week of December fifth, 1927. It discusses the difference between magnification and resolution and stresses the difficulties in obtaining clear-cut photomicrographs at high magnifications. Until comparatively recently 1500 diameters was thought to be the limit.

The ultra-violet microscope should, theoretically, give about double the resolution of one using visible light because of the shorter wave-length, but until recently this has not been the case. The paper explains a mechanical focusing method used in Bell Telephone Laboratories by which a series of photographs are taken with a change in focus of one sixteenth micron between successive exposures. A typical set of four successive exposures at 1800 magnifications is shown.

The photomicrography of steel is gone into at some length, several photographs at 3500 magnifications being shown in illustration. Special importance is laid on the preparation of samples as well as on careful focusing.

[1] *Industrial and Engineering Chemistry*, Vol. 20, No. 4, April 1928, pp. 427–429.
[2] *Mechanical Engineering*, Vol. 50, No. 3, pp. 205–212, March 1928.

Contributors to this Issue

J. H. KASLEY, Wheeling Technical College; Westinghouse Machine Company, 1898–1902; Western Electric Company, Inspection Branch, N. Y. C., 1905–07; Chief Process Inspector, New York and Hawthorne, 1908–09; Chief of Manufacturing Layout Department, 1913; Planning Engineer, 1915; Assistant Technical Superintendent, 1920; Superintendent of Manufacturing Planning, 1926; Hawthorne Works Operating Superintendent, 1927–.

F. P. HUTCHISON, B.S. in Mechanical Engineering, University of Missouri, 1916; Western Electric Company, Manufacturing and Installation Departments, 1916–. Mr. Hutchison has been engaged largely on manufacturing, planning, development, and design.

CARL R. ENGLUND, B.S. in Chemical Engineering, University of South Dakota, 1909; University of Chicago, 1910–12; professor of physics and geology, Western Maryland College, 1912–13; laboratory assistant, University of Michigan, 1913–14; Bell Telephone Laboratories, 1914–. Experimental work in radio communication is the field to which Mr. Englund has largely devoted himself since coming to the Laboratories.

J. G. FERGUSON, B.S., University of California, 1915; M.S., 1916; research assistant in physics, 1915–16; Bell Telephone Laboratories, 1917–. Mr. Ferguson's work has been in connection with the development of methods of electrical measurement.

B. W. BARTLETT, A.B., Bowdoin College, 1917; West Point, 1917–19; B.S., Massachusetts Institute of Technology, 1921; 1st Lieutenant, Engineers' Corps, 1919–22; M.A., Columbia, 1926. Mr. Bartlett was with the Bell Telephone Laboratories from 1922 to 1927, engaged in bridge development and measurement work. He returned to Bowdoin in 1927 as Professor of Physics.

O. J. ZOBEL, A.B., Ripon College, 1909; A.M., Wisconsin, 1910; Ph.D., 1914; instructor in physics, 1910–15; instructor in physics, Minnesota, 1915–16; Engineering Department, American Telephone and Telegraph Company, 1916–19; Department of Development and Research, 1919–. Mr. Zobel has made important contributions to circuit theory in branches other than the subject of wave filters.

R. V. L. HARTLEY, A.B., Utah, 1909; B.A., Oxford, 1912; B.Sc., 1913; instructor in physics, Nevada, 1909–10; Engineering Depart-

ment, Bell Telephone Laboratories, 1913–. Mr. Hartley is in charge of researches in the field of carrier current, telephone repeater, and telegraph systems.

H. A. AFFEL, S.B. in Electrical Engineering, Massachusetts Institute of Technology, 1914; research assistant in electrical engineering, 1914–16; Engineering Department and the Department of Development and Research, American Telephone and Telegraph Company, 1916–. Mr. Affel has been engaged chiefly in development work connected with carrier telephone and telegraph systems.

CHARLES S. DEMAREST, E.E., University of Minnesota, 1911; American Telephone and Telegraph Company, Engineering Department, 1911–19; Department of Development and Research, 1919–. Mr. Demarest's work has been connected with the development of equipment for toll telephone systems, including that for long distance cables, carrier and radio systems.

C. W. GREEN, B.S. in Electrical Engineering, University of Wisconsin, 1907; instructor and assistant professor, Massachusetts Institute of Technology, 1907–17; captain 1917, major 1918, U. S. Army; Bell Telephone Laboratories, 1919–. Mr. Green's work has had to do with the development of carrier telephone systems and voice frequency repeaters.

The Bell System Technical Journal

October, 1928

The Practical Application of the Fourier Integral [1]

By GEORGE A. CAMPBELL

ABSTRACT: The growing practical importance of transients and other non-periodic phenomena makes it desirable to simplify the application of the Fourier integral in particular problems of this kind and to extend the range of problems which can be solved in closed form by this method. Unless the physicist or technician is in a position to evaluate the definite integrals which occur, by mechanical means, he is usually entirely dependent upon the results obtained by the professional mathematician. To facilitate the use of the known closed form evaluations of Fourier integrals many of them have been compiled and tabulated in Table I. They are presented, however, not as definite integrals but as paired functions, one function being the coefficient for the cisoidal oscillation (or complex exponential) and the other function the reciprocally related coefficient for the unit impulse. This arrangement simplifies the table and promises to be most convenient in practical applications, since it is the coefficients of which immediate use is made, just as in the case of the Fourier series. Applications of the tabulated coefficient pairs to 85 transient problems are given, together with all necessary details, in Table II.

INTRODUCTION

THE Fourier integral and the Fourier series are alternative expressions of the Fourier theorem, the series being a limiting case of the integral and vice versa. Usually the theorem is approached from the side of the series, but there are also advantages in the approach from the integral side, which is the method followed in this paper. The generality and importance of the theorem is well expressed by Kelvin and Tait who said: ". . . Fourier's Theorem, which is not only one of the most beautiful results of modern analysis, but may be said to furnish an indispensable instrument in the treatment of nearly every recondite question in modern physics. To mention only sonorous vibrations, the propagation of electric signals along a telegraph wire, and the conduction of heat by the earth's crust, as subjects in their generality intractable without it, is to give but a feeble idea of its importance." For any real understanding of the theorem it is necessary to appreciate why it is one of the most beautiful mathematical results and why it furnishes an indispensable instrument in physics.

The Fourier integral is a most beautiful mathematical result because of the economy of means employed in obtaining a most general result.

[1] Presented September 13, 1927, in preliminary form at the International Congress of Telegraphy and Telephony in Commemoration of Volta.

One form of integral is used both to analyze and to synthesize. In both cases it is the product of the arbitrary function and the elementary sinusoidal oscillation which is integrated. This achieves the mathematical counterpart of spectrum analysis and spectrum synthesis. The functions resulting from analysis and synthesis stand in a mutually reciprocal relation.[2] They are paired with each other. Either of these functions may be assigned with an astonishing degree of arbitrariness. Singular cases being excepted, the mate function is then determined uniquely and definitely by the integral. While the sine, cosine and complex exponential are most commonly used as the elementary expansion functions, an entire class of functions present the same fundamental relations and find applications in the more recondite problems.

The Fourier integral is an indispensable instrument in connection with physical systems in which cause and effect are linearly related (so that the principle of superposition holds) because it gives at once an explicit formal solution of general problems in terms of the solution for the sinusoidal case which is often readily found. This explicit general solution makes use of two Fourier integrals, one for the spectrum analysis of the arbitrary cause and the other for the spectrum synthesis of the component sinusoidal solutions. No further consideration of the actual physical system is necessary after the elementary sinusoidal solution has been obtained. This point of view has become a part of our general background of thought.

Unfortunately the actual evaluation of specific Fourier integrals in closed form presents formidable if not insuperable difficulties. Only a small number of distinct general integrals have been evaluated in closed form in the century and more which has elapsed since the Fourier integral discovery was announced. Additions to the list of evaluated Fourier integrals can ordinarily be made only by the professional mathematician. Unless the physicist or technician is in a position to evaluate Fourier integrals by mechanical means, or is satisfied to employ infinite series or other infinite processes in place of the definite integrals, he is usually entirely dependent upon the evaluations which the professional mathematician has made in the past or is able to make for his special use. On this account, it is often desirable to so formulate practical problems that only evaluated Fourier integrals will occur. It would be well for the physicist and technician to become familiar with the Fourier integral evaluations which the professional mathematician has achieved.

[2] The fundamental importance of the Fourier integral may be associated with an analogy which exists between the integral and the imaginary unit, both considered as operators. In both cases two iterations of the operation merely change a sign and four iterations completely restore the original function.

It is the purpose of this paper to take the first steps towards the preparation of two tables, one giving the evaluations of Fourier integrals and the other giving the sinusoidal solutions for physical systems. Together they would reduce the practical application of the Fourier integral to the selection of three results from these two tables. Thus by means of the first table the arbitrary cause could be resolved into a sum of sinusoidal causes; by means of the second table the solutions for these sinusoidal causes could be supplied; and, finally, by means of the first table again, the effect of superposing these sinusoidal solutions could be shown, and thus the answer to the original problem would be given.

The preparation of the tables calls primarily for a compilation of the results already obtained by pure analysis, after which new evaluations and new solutions should be added, in so far as is possible. No attempt has yet been made to completely cover the existing literature on the subject, which extends back over one hundred years and is extensive and widely scattered. But sufficient has been done to show that the forms of the tables which are proposed are most convenient for practical application.

Paired Coefficients—Terminology

The Fourier integral theorem has been expressed in several slightly different forms to better adapt it for particular applications. It has been recognized, almost from the start, however, that the form which best combines mathematical simplicity and complete generality makes use of the exponential oscillating function $e^{i2\pi ft}$. More recently the overwhelming advantage of using this oscillating function in the discussion of sinusoidal oscillatory systems has been generally recognized. It is, therefore, plain that this oscillating function should be adopted as the basic oscillation for both of the proposed tables. A name for this oscillation, associating it with sines and cosines, rather than with the real exponential function, seems desirable. The abbreviation cis x for $(\cos x + i \sin x)$ suggests that we name this function a cis or a cisoidal oscillation. This term is tentatively employed in this paper. The notation $\text{cis}(2\pi ft)$ is also employed where it is desired to use an expression which is essentially one-valued, which avoids the use of exponentials, or which suggests periodic oscillations by its connection with cosine and sine.[3]

[3] Since the cisoidal oscillation is simply a uniform rotation at unit distance about the origin in the complex plane, it may be desirable to try some compact notation which directly suggests this rotation; for example, $\text{ru}(ft)$, 1^{ft}, i^{4ft} might be defined as the complex quantity obtained by rotating unity through ft complete turns or $4ft$ quadrants.

In a table of Fourier integrals, every integral expression would then contain, in addition to the arbitrary function $F(f)$, the same oscillating function $\text{cis}(2\pi ft)$, the same integral sign with limits $-\infty$, $+\infty$ and the same differential df. To repeat any such group of a dozen characters in each of several hundred entries seems quite unnecessary. It is, therefore, proposed merely to tabulate the arbitrary function $F(f)$ and the value $G(t)$ for the evaluated integral expressed as a function of the time. The table is thereby reduced to two parallel columns of associated functions, one of which is employed as the coefficient of the elementary cisoidal function while the other is a function of the independent time variable. The table would, however, be more symmetrical if both of the associated functions could be regarded as coefficients of an elementary function. This may be done by introducing the unit impulse as an elementary function, the impulse occurring at the epoch g at which instant it presents a unit area whereas its value is zero for all time before and all time after the epoch g. This is an essentially singular function and to recognize this fact it will be designated by $\mathbb{S}_0(t-g)$ which is intended to emphasize the singularity. The time function may now be replaced in the table by the same function of the parameter g, since the time function $G(t)$ is equal to the integral with respect to g between infinite limits of the product $G(g)\mathbb{S}_0(t-g)$.

The table of Fourier integrals has now become also a table of paired coefficient functions. This means that if the coefficient $F(f)$ is employed with the cisoid, and the coefficient $G(g)$ is employed with the unit impulse, and both products are summed for the entire infinite range of their parameters f and g, the same identical resulting time function is obtained.[4] Taken in connection with their respective elementary functions, the two associated coefficient functions are, therefore, equivalent, alternative ways of representing a particular time function. This is the fundamental geometrical or physical point of view which is needed in connection with the practical application of the Fourier integral theorem. For this reason the table has been headed a table of Paired Coefficients; as explained above, however, it may equally well be considered to be a table of Fourier Integrals.

There is another fundamental reason for placing both of the functions $F(f)$ and $G(g)$ on the same footing as coefficients. It is this:

[4] The use of frequency and epoch as the two parametric variables gives us many symmetrical formulas where, if the radian frequency were employed, an unsymmetrical 2π would occur. In practical applications the frequency of the coefficient pairs becomes the frequency which is ordinarily employed in acoustics, in music and in commercial alternating currents. The basic unit for frequency is the reciprocal second; the unit for epoch is the second.

Fourier's fundamental discovery was that the two functions may be transposed in the Fourier integral if the sign of one of the parameters is reversed. Thus, either one of the two functions constituting any coefficient pair may be taken as the coefficient of the cisoidal oscillation, provided only that the proper sign is given the epoch parameter occurring in the other function. For this reason also both functions are thus quite properly regarded as coefficients.

It is found convenient to call each coefficient of a coefficient pair the mate of the other coefficient, pair and mate being employed just as in the case of gloves. To find the mate of a glove, it is necessary to know all about the given glove including the fact as to whether it is the right or the left one of the pair. In the same way, to find the mate of a coefficient function, it is necessary to know not only the form of the function, but, in addition, whether its variable is the frequency or the epoch. The notation $\mathfrak{Mate}\,G(g)$, $\mathfrak{Mate}\,F(f)$ will be employed to indicate the mate of the particular coefficient $G(g)$, $F(f)$.

We have now defined and explained the proposed terminology for use in the practical application of the Fourier integral theorem. Before proceeding to practical applications, it is desirable to become familiar with these coefficient pairs considered in their own right. We may well begin by reminding the reader of the dissimilarity between the elementary oscillations.

The Two Elementary Functions Contrasted

The dissimilarity between the two elementary functions of the time, the cisoidal oscillation $\operatorname{cis}(2\pi ft)$ and the unit impulse $\mathfrak{S}_0(t-g)$ is most striking. This is clearly shown by the wire models of Fig. 1 where each function is depicted for five values of its parameter. For the value zero the cisoidal oscillation degenerates into an infinite straight line parallel to the time axis and cutting the real axis at $x = 1$. For the same value zero of its parameter g, the unit impulse is zero everywhere except at the origin where it has a vanishingly narrow loop extending to $x = +\infty$.

For other values of the parameter, the cisoidal oscillation is always an infinite cylindrical helix, centered on the time axis, and passing through the point $x = 1$, while the infinite loop of the impulse function is displaced unchanged along the time axis to $t = g$. For positive values of the parameter f, the cisoidal oscillation is a right-handed helix, with pitch equal to f^{-1}, and thus decreasing as f increases. For negative values of f, the pitch is the same but the helix is left-handed.

Both functions have essential singularities, which are quite dif-

ferent both in character and in location. For the cisoidal oscillation the singularity is always located at $t = \infty$; for the impulse the singularity is at $t = g$.

The fundamental differences between the two elementary time functions adapt them for different uses. It is desirable to be in a position to employ first one and then the other, shifting from one to the other without any trouble or delay, so that at each step of a problem the elementary function best suited for use may be employed.

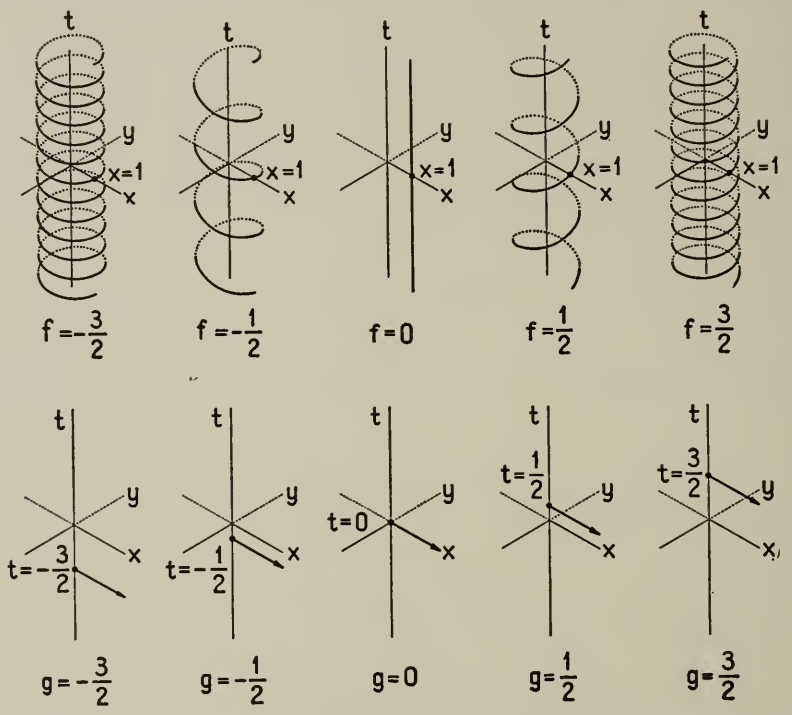


Fig. 1—Wire models of cisoidal oscillations cis $(2\pi ft)$ (above) and of unit impulses $\mathfrak{S}_0(t - g)$ (below) for the particular values 0, $\pm 1/2$, $\pm 3/2$, of the parameters f and g.

For this we require only an adequate table of pairs and a certain familiarity in the use of the pairs. It is desirable to acquire the habit of thinking of the coefficients of a pair as alternative representations of a curve.

The Use of Table I for Obtaining Coefficient Pairs[5]

The table is divided into nine parts. In Part 1 are given the general processes for deriving any coefficient mate; but such processes are to

[5] Five other closely related uses may be made of Table I as explained in the first footnote to that table. Operational expressions are brought within the scope of the table by substituting for the operator $p = d/dg$ the particular value $i2\pi f$, other possible interpretations of the operator, if any, being ignored.

be employed only when it is necessary to start from first principles. All mates which have once been determined may be taken from the latter sections of the table with a great saving of time and energy. Part 2 of the table shows the elementary transformations and combinations of pairs; these theorems may be employed either to extend a given table of coefficient pairs, or to cover a given group of coefficients with a shorter table of specific pairs. It is assumed that anyone desiring to make serious use of the table will first become familiar with these elementary combinations and transformations; even the simple addition, factor and transposition theorems (201), (205), (217) are most useful.

Part 3 of the table contains seven pairs, which are called key pairs because all specific pairs listed in the entire table may apparently be derived from them by specialization or by passing to a limit after any necessary use has been made of the elementary combinations and transformations of Part 2, amplified, as indicated, by the removal of certain unnecessary restrictions to real quantities. If an assigned coefficient is not included in Part 3 as thus generalized, then this coefficient cannot be found anywhere in Table I. Part 3, therefore, serves the useful purpose of giving a bird's-eye view of the entire table. The seven pairs are presumably redundant as they stand.

For applicational purposes it is most desirable to have a table which lists the precise pair required; many special cases which have been used in practical applications may be found in Parts 4–9 of Table I which constitute a short classified list of particular cases. It is important to remember that a given coefficient should be looked for on the other side of the table if it is not found on its own side since all pairs are transposable by (217) or (218). In the tables as they stand, some pairs have been transposed, but this is not true in the majority of cases.

Whenever an infinite process is to be employed, such as infinite series, integration or differentiation, the permissibility of the process is a question which must be answered for the particular case in hand; the formal result given in Table I may break down, for example, if either the original or the transformed pair is a singular pair. This general warning necessarily applies to every part of the subject of coefficient pairs just because it is a part of the general subject of mathematical analysis.

It is intended that the statement of each pair in the entire table shall eventually include every limitation and every warning which the mathematical sponsors for that pair would consider necessary to guarantee its safe use by anyone understanding the fundamental

nature of coefficient pairs. A beginning has been made by specifying the branches of multiple-valued functions and the method of approaching limits. When this has been fully carried out, any pair may be taken from the table and used without the least concern as to the analytical methods by which the validity of the pairs has been established. Thus the finished table will make possible a complete separation of the analytical evaluation of all known Fourier integrals from their practical applications.

Having now explained, in a general way, the use of Table I, it will be useful to consider in detail a limited number of the pairs which are of special practical interest.

General Processes for Deriving the Mate

The table is naturally headed by the two fundamental Fourier integrals (101), (102) because of their intrinsic importance as explicit and implicit definitions of coefficient mates. The chief purpose of the table, however, is to make it possible for the technical man to make the fullest use of coefficient pairs without concerning himself at all as to the analytical work of evaluating either of these Fourier integrals. Pairs (101) and (102) are thus intended to serve mainly as definitions for the pairs which follow.

The statement has been made that essentially only one Fourier integral has been evaluated by determining the indefinite integral and substituting the integration limits. Whether or not this is precisely true, the statement does illustrate the fact that the formulation of the Fourier integral does not in itself suggest a practical finite analytical process for the actual evaluation of the definite integral. No such system of evaluating definite integrals is known. Writing down the Fourier integral amounts to little more than definitely formulating a question.

If the coefficient $F(f)$ is expanded as a finite or infinite series in powers of f (or p), the mate is given by pair (106*), and this involves a finite or infinite series of essentially singular functions which are further considered below in connection with Fig. 3. If a series expansion of $F(f)$ is made in terms of any functions of f for which the mates are known, there is a corresponding series for the mate. Some of these pairs are shown as (104*)–(112). The possibility of the formal infinite expansion does not necessarily imply the convergence of the series in the case of coefficient pairs any more than in other general developments.

The technical man is not ordinarily a master of infinite series, definite integrals or other infinite processes. It is, therefore, highly

desirable to give him coefficient pairs which are in closed form, that is, involve only a finite number of operations with known functions. Accordingly, the portion of the table expressible in closed form has seemed to be the part which should be developed first. Specific pairs requiring infinite series for their expression have not been included in this preliminary draft of Table I. The omission of these series and of other infinite processes does not signify any failure to appreciate their importance. It is intended to include specific infinite series later.

The Elementary Transformations of Coefficient Pairs

The simple addition theorem (201) is of the greatest practical importance. The summation may include any number of pairs; they may be quite unrelated, or they may be the successive terms of power expansions as shown in (106*)–(111*). Next to the addition theorem we may place the multiplication theorem (202) or (203), special cases of which are of great practical importance. Among these special cases are (206)–(211) where any coefficient is multiplied or divided by its parameter or by a cisoidal oscillation of its parameter.

Any real linear substitution for the frequency and epoch parameters is made possible by the simple transformations (205)–(207), (214). The generalization of these transformations by the removal of the restriction to real numbers is allowable in important cases as is indicated by the parameters shown in square brackets with each pair of Part 3.

The differentiation and integration of coefficients with respect to the frequency, epoch or other parameter give the important transformations (208)–(213).

Some of the simple transformations continue to yield new results when they are repeated any number of times or when several transformations are combined in sequence. Pairs (216), (218)–(222) are examples of such combinations. All pairs in Parts 4–9 of this table may apparently be derived from the seven key pairs of Part 3 by means of these transformations employing complex parameters as indicated in Part 3, and passage to a limit in certain cases.

The resolution of pairs into the four types of i^n-multiple pairs, as shown by pairs (223)–(225), throws considerable light on the nature of coefficient pairs.

Some of the elementary properties of pairs are expressed in words as follows:

Elementary Properties of Pairs

(1) The sum or difference of pairs is a pair. Cf. pair (201).

(2) Any constant multiple of a pair is also a pair. Cf. pair (204).

(3) Any linear combination of pairs is also a pair. Cf. pairs (201), (204).

(4) The odd and even parts of every pair are also pairs.

(5) If both coefficients of a pair are real, both are even.

(6) If a pair has one real and one pure imaginary coefficient, both are odd.

(7) If a coefficient is even and real, its mate is also even and real.

(8) If a coefficient is odd and real, its mate is odd and pure imaginary, and vice versa.

(9) If a coefficient is real, its mate has conjugate values for opposite values of its parameter and conversely. Cf. pair (216).

(10) The conjugates of the coefficients of a pair are also a pair provided the sign of either frequency f or epoch g is reversed. Cf. pair (215).

(11) A pair with the signs of both frequency f and epoch g reversed is also a pair. Cf. pair (214).

(12) The parameter of either coefficient may be multiplied by a positive real constant provided the other parameter and coefficient are each divided by the same constant. Cf. pair (205).

(13) Coefficients of a pair may be interchanged if, when interchanging the parameters, the sign of one parameter, either f or g, is reversed. Cf. pair (217).

(14) Any pair may be resolved uniquely into the sum of four pairs by pairing together: the even, real parts; the even, imaginary parts; the odd, real part of each coefficient with the odd, imaginary part of the other coefficient.

(15) A pair may have the form $(F(f), \lambda F(g))$ where the multiplier λ is constant, if and only if λ has one of the four unit values $(1, i, -1, -i)$. Such a pair is called an i^n-multiple pair. Cf. pair (223).

(16) Any i^n-multiple pair has both coefficients odd or even according as n is odd or even.

(17) Any i^n-multiple pair with complex coefficients may be resolved into two i^n-multiple pairs with coefficients which are real or pure imaginary.

(18) The coefficients of any two i^n-multiple pairs are orthogonal if the i^n multipliers are different.

(19) The coefficients of any four i^n-multiple pairs with different i^n multipliers are linearly independent.

(20) Any pair may be resolved uniquely into the sum of four i^n-multiple pairs; i.e., pairs of the form $F_n(f)$, $i^n F_n(g)$. Cf. pair (224).

(21) Any pair may be resolved uniquely into the sum of eight i^n-multiple pairs where $F_n(f)$ is real or pure imaginary. Cf. pair (225)

Pairs Based on the Normal Error Law

The identical pair (703), $\exp(-\pi f^2)$, $\exp(-\pi g^2)$, is one of the simplest pairs and may well serve as the starting point in the consideration of specific coefficient pairs. Each coefficient is the broad impulse of the normal error law. It is remarkable that identical coefficients of this simple form should produce the same identical function when associated with either the cisoidal oscillation or the very different unit impulse.

If the differential transformation (222), taking the upper signs, is applied to the normal error law pair (703), the infinite series of ϕ_n pairs (702) is obtained. Of these derived pairs, the first eight are written out as pairs (704)–(711). The cisoidal coefficients are alternately even and odd functions which oscillate in the neighborhood of the origin, each successive coefficient having an added half oscillation. The ϕ_n pair has $(n + 1)$ half oscillations. Beyond these oscillations, every coefficient in the infinite sequence decreases rapidly and asymptotically to zero in both directions. The mates of these cisoidal coefficients are identically the same except for a constant coefficient which is i^n and thus goes cyclically through the four values, 1, i, -1, $-i$.

The $\phi_n(x)$ functions are shown by Fig. 2. They are essentially the parabolic cylinder functions of order n. These coefficients may be used for the expansion of every function which, with its first two derivatives, is continuous for all positive and negative values of the variable and for which a certain integral exists. This expansion is known as the Gram-Charlier series, which appears in pair (112).

Starting again with the normal law of error pair (703) in the form (701) and setting $\rho = \frac{1}{4}\beta/\pi$, and applying the differential transformation (208) repeatedly, we obtain the infinite sequence of pairs (713) of which the first five are listed as pairs (714)–(718). The cisoidal coefficients are the successive integral powers of p multiplied by the normal error exponential. The impulse coefficients are essentially the ϕ_n functions multiplied by the normal error exponentials. These pair, are plotted in Fig. 3 for the special case $\beta = a^2 = 1$.

Both of the infinite series of pairs derived from the error function and shown in Figs. 2 and 3 are regular throughout, are nowhere infinites and vanish at infinity.

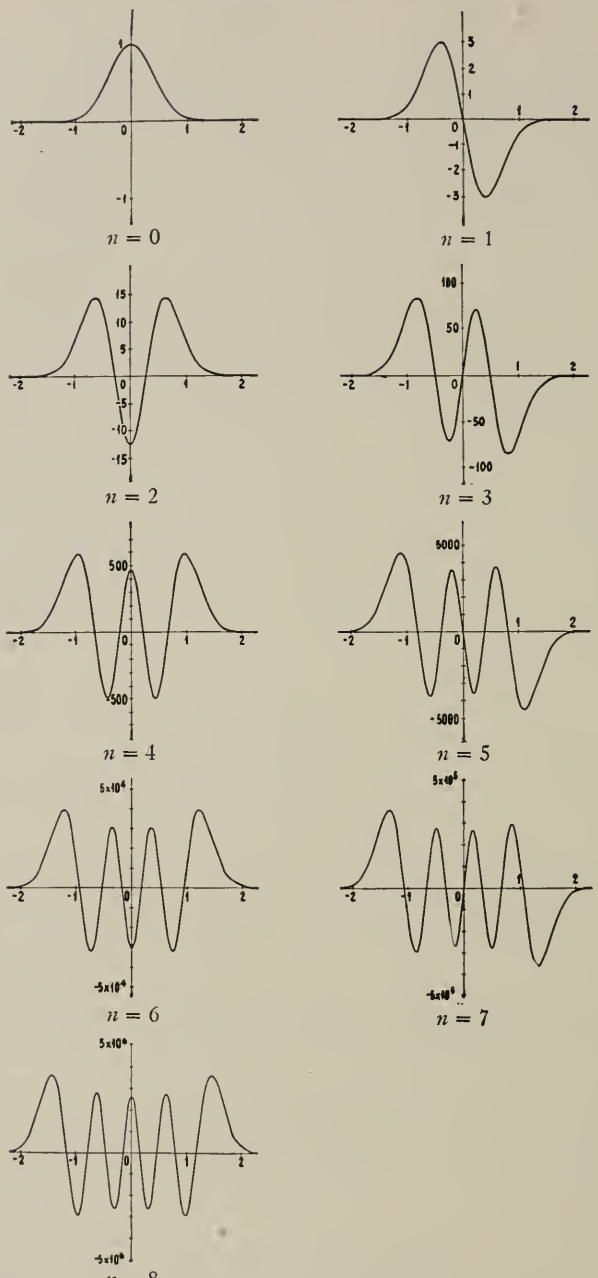


Fig. 2—Curves showing the $\phi_n(x)$ functions for $n = 0, 1, 2, \cdots, 8$
$\phi_n(x) = \exp(\pi x^2) D_x{}^n \exp(-2\pi x^2)$.

Essentially Singular Pairs for Integral Powers of the Parameter

If in Fig. 3, with the value of n held fixed, we allow a to approach the limit 0, the cisoidal coefficient becomes p^n and the impulse coefficient, which is compressed horizontally towards the origin and expanded vertically, with corresponding areas increasing as a^{-n}, ultimately vanishes everywhere except at the origin where it acquires an essential oscillating singular point. At the limit, then, a singular pair is obtained; it will be designated as p^n, $\mathbb{S}_n(g)$. $\mathbb{S}_n(g)$ is characterized by having all of its moments about the origin vanish except the nth moment, which is equal to $(-1)^n n!$ The dotted graphs on the left of Fig. 3 show p^n to the scales indicated. The curves on the right show $\mathbb{S}_n(g)$ provided we assume that the horizontal scale is increased with a and the vertical scale increased inversely with a^{n+1} as a approaches the limit 0. Fig. 3 thus serves to picture the essentially singular function $\mathbb{S}_n(g)$. That is, it is sufficient if the coefficient maintains this form while proceeding to the limit. This form is, however, not essential. It is necessary only that the method of approach to the limit give the same set of moments.

An alternative way of deriving the mate for the positive integral powers p^n is by means of a linear combination of $(n+1)$ pairs of the form of (603) with parameters equal to a, $2a$, $3a$, $\cdots$, $(n+1)a$, respectively, so that the first term in the power series expansion of the cisoidal coefficient is p^n. The corresponding impulse coefficient is a succession of $(n+1)$ bands, each of width a, the first band beginning at epoch zero, the heights of the successive bands being equal to the binomial coefficients for power n divided by a^{n+1} but alternately positive and negative. The mth moment of this impulse coefficient is 0 for $m < n$, equal to $(-1)^n n!$ for $m = n$, and proportional to a^{m-n} for $m > n$. Upon allowing a to approach zero, the cisoidal coefficient approaches p^n, and the impulse coefficient approaches $\mathbb{S}_n(g)$, since in the limit the same set of moments is obtained as was found above to characterize the nth singularity function. This is pair (402*).

The special cases for $n = 0$, 1 are of most frequent occurrence. They are pairs (403*), (404*). $\mathbb{S}_0$ is the unit impulse since its 0th moment equals unity; $\mathbb{S}_1$ is the doublet with the moment -1 since its first moment is -1. $\mathbb{S}_1$ and all higher order singular functions are included in the series coefficients of (104*), (106*).

Fig. 3 may be extended upward step by step from the normal error law pair by dividing by p on the left and integrating with respect to g on the right. At each step a constant of integration is introduced. The first two pairs thus obtained are pairs (725*) and (726*). Choos-

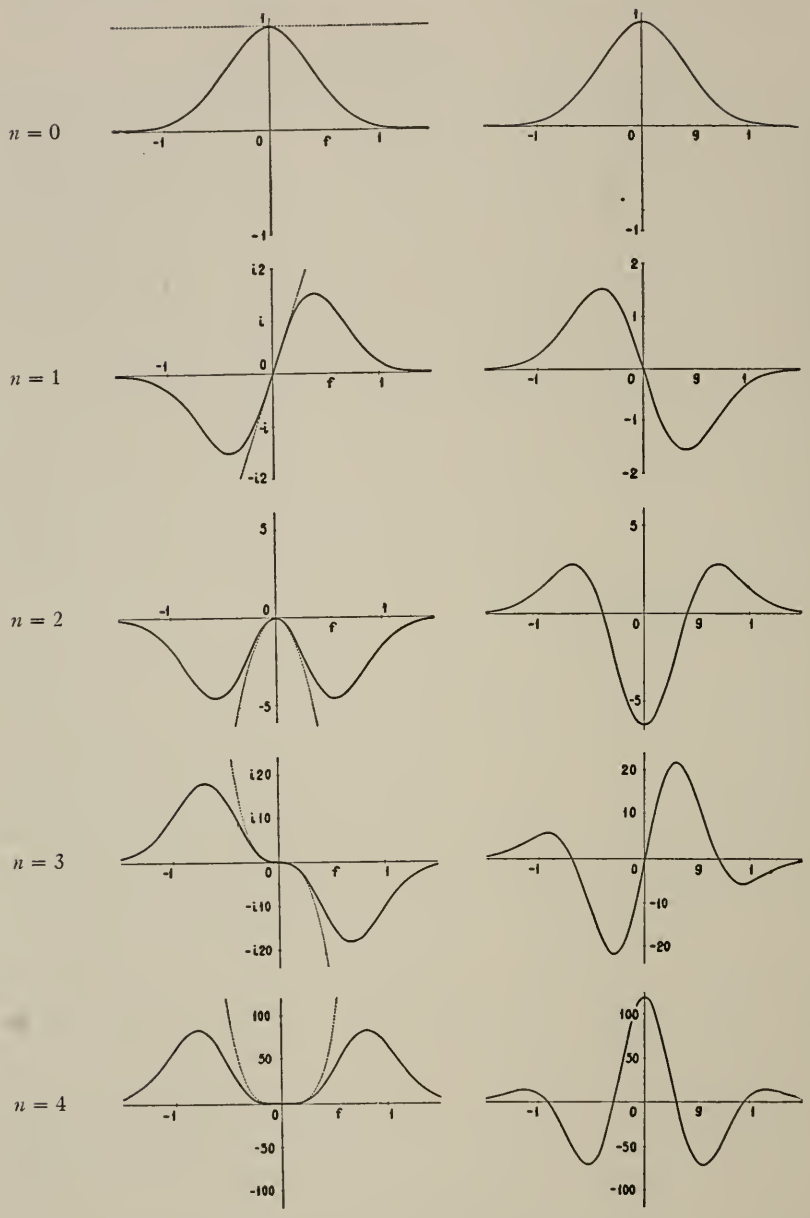


Fig. 3—Graphs for the family of pairs $p^n \exp(-\pi a^2 f^2)$ $a^{-1} D_g{}^n \exp(-\pi g^2/a^2)$. The heavy curves show the cases $a = 1$, $n = 0, 1, 2, 3, 4$; the dotted curves on the left are for the same values of n but for the limit $a \to 0$. On the right the curves apply for any value of a if the horizontal and vertical scales are multiplied by a and a^{-n-1} respectively.

ing the integration constants so as to make the impulse coefficients alternately odd and even, these two pairs are as shown in Fig. 4. If we now allow a to approach the limit zero, a new series of pairs is obtained of which the first two pairs are shown dotted in Fig. 4 for the particular choice of integration constants there made. The general limiting pair is designated as p^{-n}, $\mathbf{S}_{-n}(g)$ and it is shown with its n arbitrary parameters $\lambda_1, \lambda_2, \cdots, \lambda_n$ as pair (410*). In some ways it is simpler to derive the limiting pair for negative integral powers of p from rational functions of p, which may be accomplished as shown by pair (411*). Special cases are shown by pairs (408*), (409*), (415*), (416*).

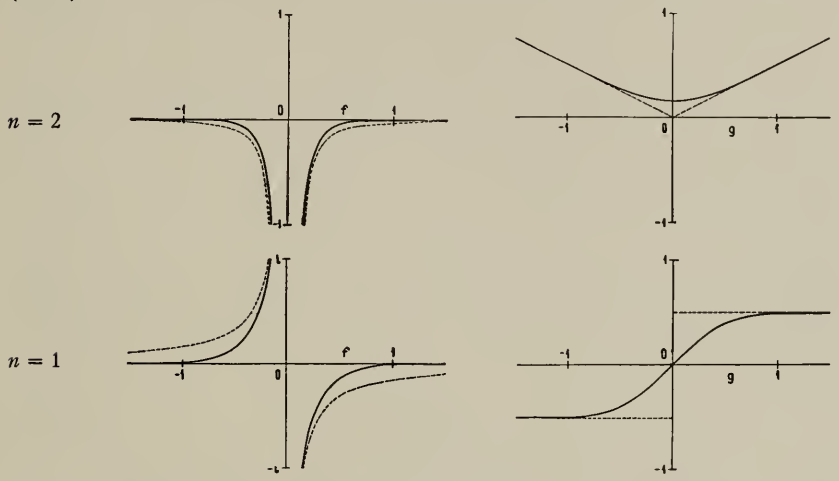


Fig. 4—Graphs for the family of pairs $p^{-n}\exp(-\pi a^2 f^2)$, $a^{-1}D_g{}^{-n}\exp(-\pi g^2/a^2)$, with the integration constants chosen so as to make the impulse coefficients alternately odd and even. The heavy curves show the cases $a = 1$, $n = 1, 2$; the dotted curves show the limit $a \to 0$, $n = 1, 2$.

The first of the series $\mathbf{S}_{-1}(g)$ is a unit step at epoch 0 from a constant value $\lambda - \frac{1}{2}$ for all negative epochs to the constant value $\lambda + \frac{1}{2}$ for all positive epochs. The constant λ may have any value; this is a singular case marked by the failure of the general rule that the choice of the cisoidal coefficient uniquely determines the impulse coefficient. This means that in any well set problem some other condition determines the value of the constant λ. In some problems, for example, it is necessary that the epoch coefficient be an odd function, and then λ vanishes. In other problems where either the epoch function must be zero for all negative epochs or on the other hand the p occurring in the cisoidal coefficient is actually the limit of $p + a$ as a approaches zero through positive values, the constant λ equals $\frac{1}{2}$. This limiting condition may arise if we assume that resistance may be ignored, as a first approxima-

tion, in studying actual systems which necessarily involve at least a small amount of dissipation.

The mates of positive and negative integral powers of p, including the zero power, cannot be derived directly and definitely from the Fourier integral (101) without the specification of an additional passage to a limit. Such pairs therefore differ essentially from the great body of regular pairs where the choice of one coefficient completely determines the mate. In order to permanently ear-mark these limiting pairs, their serial numbers in Table I bear a star. These pairs may be thought of as lying on the periphery of the great domain which includes the totality of regular pairs.

Identical Mates and Other Simply Related Mates

Since one of the coefficients of a pair may be assigned quite arbitrarily, this choice allows us, if we so elect, to specify some relation between the two coefficients of a pair. We might specify that a linear combination $\lambda F_j(x) + \mu G_j(x)$ of the two coefficients of a pair both taken with the parameter x is to equal an arbitrary function $F(x)$. The pair (F_j, G_j) is then uniquely determined, unless $\lambda + i^n\mu = 0$, being equal to pair (224) after each F_n has been divided by $\lambda + i^n\mu$. Again if it is specified that one coefficient is to be the reciprocal of the other, a possible solution is pair (760).

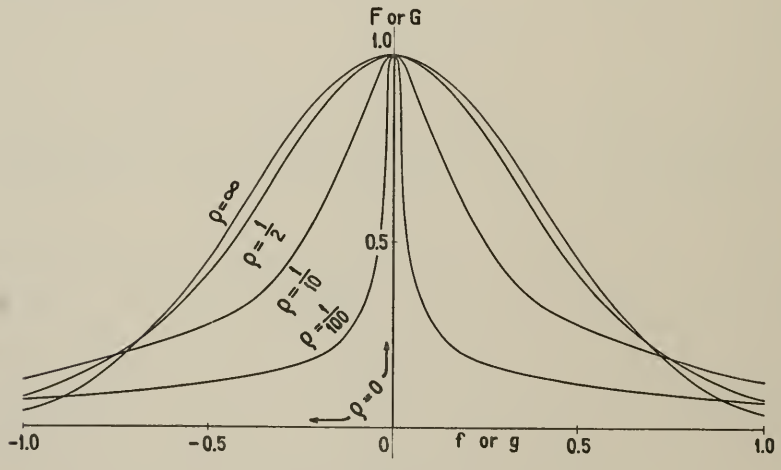


Fig. 5—Identical coefficient pairs of the form
$(1 + x^2/\rho^2)^{-\frac{1}{4}} K_{\frac{1}{4}}(2\pi\rho^2\sqrt{1 + x^2/\rho^2})/K_{\frac{1}{4}}(2\pi\rho^2)$, $x = f$ or g.

The condition that the mates shall be identically the same function of their parametric variables f and g is of special interest. In addition to the identical pairs shown on Fig. 2, $n = 0, 4, 8$, the table contains a number of identical pairs including (523), (625), (712), (761), (916).

The identical pair (916) divided by its value at the origin is shown in Fig. 5 for different real values of its parameter ρ. For $\rho = +\infty$, the curve is of the $\exp(-\pi x^2)$ or normal law of error form, and is identical pair (703). For $\rho = \frac{1}{2}$, the reciprocal hyperbolic cosine identical pair (625) is shown correctly within the width of the line, this being apparently a mere coincidence since pair (916) does not include it as a special case. Finally, for $\rho = 0$, the limiting curve coincides with the horizontal axis taken together with unit length of the positive vertical axis. This represents pair (523) divided by its value at the origin, which is infinite. The point to be especially noted is that the area under every curve of the family illustrated by Fig. 5 is the same and equal to unity. This must hold for the limit $\rho = 0$, when the curve encloses no area within a finite distance of the origin.

The identical pair $|f^{-\frac{1}{2}}|$, $|g^{-\frac{1}{2}}|$ is of great simplicity and it occupies a central position among algebraic pairs. Starting with the minus one-half power of the parameters in both coefficients, any increase in the power of one parameter requires an equal decrease in the power of the other parameter as is illustrated, for example, by pairs (502*), (516*), (524).

It is not permissible to specify any relation whatsoever between the two coefficients of a pair; for example, no pair exists for which one coefficient is twice the other. As stated above, the only multiples permissible are the four units 1, i, -1, $-i$. For each of these four cases there are an infinite number of solutions. These solutions satisfy the integral equations given in the foot-note to pair (223).

Practical Applications of Coefficient Pairs

Fourier gave the first comprehensive method of finding the solution for transients. His method involves three steps: viz.,

I. Spectrum analysis of the cause among all frequencies.
II. Solution for all frequencies.
III. Spectrum synthesis of the effects for all frequencies.

Fourier thus substituted three problems for one. With a table of Fourier coefficient pairs, these three steps may be made as follows:

I. Find the mate of the cause considered as an impulse coefficient.
II. Multiply this mate by the admittance for the system.
III. Find the mate of this product considered as a cisoidal coefficient.

These three steps define a perfectly definite result, since every arbitrarily chosen coefficient has a mate which is unique and determinate, or may be made so by the specification of some suitable passage to a limit.

The use of a table of pairs may also be stated in another and somewhat more general way as follows:

For any system where the principle of superposition holds, any cause $C(t)$, its effect $E(t)$ and the corresponding admittance $Y(f)$ are connected by a relation which may be written in any one of three ways which explicitly express each of the three quantities in terms of the remaining two, as follows:

$$E(g) = \partial\mathfrak{l}[Y(f)\partial\mathfrak{l}C(g)],$$

$$C(g) = \partial\mathfrak{l}\left[\frac{\partial\mathfrak{l}E(g)}{Y(f)}\right],$$

$$Y(f) = \frac{\partial\mathfrak{l}E(g)}{\partial\mathfrak{l}C(g)},$$

where $\partial\mathfrak{l}$ is read "mate of."

The use of coefficient pairs may be most simply illustrated by reference to Figs. 3 and 4, in connection with the problem of finding transient currents through a perfect condenser of unit capacity due to impressed electromotive forces shown by each of the seven curves on the right considered as functions of the time. Any curve on the right being the cause, the next curve below it is the effect, considering Fig. 4 to be placed above Fig. 3. In the solution the first step is to find the mate of the curve on the right. This is the curve on the left. This mate is then to be multiplied by the admittance of the system which is p for a unit condenser. Reference to the titles of the figures shows that this product is given by the next lower curve on the left. To find the mate of this last curve is the third step in the solution and for this it is merely necessary to go to the curve on the right. The three steps then take us from any curve on the right to the next curve below it. Figs. 3 and 4, taken together, are a section of an infinite sequence of pairs which illustrate an infinite number of possible transients in a perfect condenser of unit capacity.

If, on the other hand, the system consisted of a perfect reactance coil of unit inductance and the impressed cause was again shown by any curve on the right, the effect would be shown by the next higher curve, assuming that the initial current at the beginning of time was that shown by the extreme left of the upper curve. Thus, when the cause is oscillating, there is one less half oscillation in the effect than in the cause. This is for an inductance. For a condenser, conditions are reversed; the effect has one more half oscillation than the cause.

The scales of Figs. 3 and 4 may be changed to correspond to any value of a, the parameter which appears in the coefficients of the pairs.

At the limit $a = 0$, the cause and effect would be the singular $\mathfrak{S}_n$ or $\mathfrak{S}_{-n}$ functions.

The curves on the right for $n = 0$ of Fig. 3 and $n = 1$ of Fig. 4 show that at the limit $a = 0$ a unit step in the voltage produces a unit impulse in the current through a unit condenser; on the other hand, a unit impulse applied to a unit inductance gives a current which is a unit step.

The curves of Fig. 2 may be used to furnish another illustration of the use of coefficient pairs, in connection with the problem of finding networks in which assigned transient currents will be produced by assigned impressed electromotive forces. Any curve n being the assumed cause and the next curve $(n + 1)$ the assumed effect, the required admittance is $\phi_{n+1}(f)/[i\phi_n(f)]$. This admittance is presented by a ladder network of $(n + 1)$ elements: perfect inductance coils in the series arms, perfect condensers in the shunt arms, the ladder starting with a shunt condenser, the values of the shunt capacities being equal to 2, $2n(n-1)^{-1}$, $2n(n-1)^{-1}(n-2)(n-3)^{-1}$, etc., and the values of the series inductances being equal to $(2\pi n)^{-1}$, $(2\pi n)^{-1}(n-1)(n-2)^{-1}$, etc. In verifying the solution of this problem, it is to be noticed that the mates of the curves n and $(n + 1)$, regarded as impulse coefficients, are the same curves multiplied by i^{-n} and $i^{-(n+1)}$; the quotient of the latter mate divided by the former mate is the admittance of the network as given above.

On the other hand, any curve $(n + 1)$ being the cause, the curve n is the effect in the reciprocally related ladder network of $(n + 1)$ elements, starting with a series reactance coil, the values of the series inductances being equal to 2, $2n(n-1)^{-1}$, $2n(n-1)^{-1}(n-2)(n-3)^{-1}$, etc., and the values of the shunt capacities being equal to $(2\pi n)^{-1}$, $(2\pi n)^{-1}(n-1)(n-2)^{-1}$, etc.

Practical Applications of Coefficient Pairs in Table II.

In general, each of the three subsidiary problems employed by Fourier is unsolvable in closed form. In a strictly limited number of cases, however, all three problems have been solved and the final transient solution obtained. These solutions should be cherished and collected for ready reference. It is a needless waste of time to repeat the analytical work each time a solution is required. Except for a few special cases lying outside of the scope of the table, all practical applications of closed form coefficient pairs which were found in a preliminary search are included in the transient solutions of Table II. As it stands, the table is far from a complete list of closed form solutions, but it contains many important solutions and serves to illustrate the use of Table I. Table II contains 39 admittances, with references to 39 systems which serve to illustrate the occurrence of these admit-

tances. In the third, fourth and fifth columns, 85 transient solutions are given of which 39 are for the unit impulse, 30 for the unit step, and 16 for the suddenly applied cisoid.

The causes producing the transients in Table II are but three in number: the unit impulse, the unit step, and the suddenly applied cisoid; and the mates for these causes are unity, p^{-1} and $(p - p_0)^{-1}$ as is shown by pairs (403*), (415*) and (440*). Multiplying these three mates by the admittances and taking the mates of the products, we have the effects, as is stated in the headings of the last three columns of the table.

To illustrate in detail the steps involved in finding a transient effect with the aid of Table I, consider system No. 14 of Table II with the cause equal to the unit step $\mathbf{S}_{-1}(t)$, $\lambda = \frac{1}{2}$. The mate of the unit step is p^{-1} by pair (415*). Multiplying this by $Y(f)$ as given in the second column of Table II, we have $up^{-\frac{1}{2}}(1 + \sqrt{p/\lambda})^{-1}$ for the cisoidal coefficient. By pair (551) the mate of this is $u\sqrt{\lambda}\exp(\lambda g)\operatorname{erfc}\sqrt{\lambda g}$, $0 < g$. Substituting for g the actual variable t, we have the transient solution as given in the fourth column and fourteenth row of Table II.

This simple example fully illustrates the three essential steps in finding any transient effect when the admittance and pairs are known. In this example the effect was considered to be the unknown. If either the cause or the admittance were the unknown, the same pairs would be involved but the two coefficients in a pair would be used in the reversed sequence in all but one instance.

There are still 32 squares of Table II left blank. It would be a simple matter to place series solutions or integral solutions in each of these squares. Thus if the impulse transient of column 3 is known, the other two transients are given at once in integral form by pairs (210) and (219); if the unit step transient of column 4 is known, the suddenly applied cisoidal transient is written immediately in integral form by the use of pair (220). The real problem is, however, either to find closed form solutions in terms of known functions or to show that this is impossible. When the failure of known functions has been established, we should next consider the choice of new functions so defined as to throw as much light as possible on the new solutions.

Table II may be regarded as another table of coefficient pairs. Column 2 contains cisoidal coefficients; column 3, the mates of these coefficients; column 4, the mates of these coefficients when multiplied by p^{-1}; and column 5, the mates of these coefficients when multiplied by $(p - p_0)^{-1}$. The corresponding pair in Table I is referred to in the lower left-hand corner of each square by its serial number. In a few cases, two or three pairs are referred to and there it is necessary

to add the Table I pairs together or, in the case of systems 37–39, to apply the two pairs in sequence. In Table II, the customary physical notation is adhered to because it is often of long standing and this necessitates some change in notation when comparing pairs in the two tables.

Summary and Conclusions

Many practical applications of the Fourier integral have been simplified by the compilation of Tables I and II, which give coefficient pairs, admittances and transient solutions.

Minor changes in nomenclature and point of view have been introduced, all with the idea of simplifying the practical application of the Fourier integral, in the following ways:

(1) Using the cisoidal oscillation and the unit impulse side by side as alternative elementary expansion functions.

(2) Focusing attention upon coefficient pairs for these two elementary functions, both coefficients of a pair representing the resolution of the same arbitrary function.

(3) Using the frequency and epoch as the parametric variables, in place of the customary radian frequency and independent time variable.

(4) Employing as a coefficient any real or complex arbitrary function which may be practically useful by regarding it, where necessary, as a limit approached through coefficients which form regular pairs.

(5) Introducing the $\mathfrak{S}_n(g)$ functions having an essential oscillating singularity at the origin which mate with p^n, the positive integral powers of p.

(6) Using a notation which greatly reduces the number of occasions for employing the integral symbol in applications of the Fourier theorem.

Having established the inclusiveness and practical utility of the proposed coefficient pair method of applying the Fourier integral, we are now planning to critically verify the tables and make them as complete as is feasible. It is proposed to include eventually such references to the literature as may add to the interest of the tables. The contributions of integral equations and of the operational method to the present subject will also be incorporated in the tables. The preparation of similar tables for other elementary expansion functions, such as Bessel functions, is also a possibility. A comprehensive table might be made which would include in parallel columns the coefficient functions for a large number of elementary expansion functions, thus giving at once many alternative ways of representing particular time

functions. This would make it possible to shift without trouble from any one expansion to any other expansion of the tabulation.

I am under great obligations to my colleagues for their contributions towards the preparation of this paper. I shall be grateful to any person who will call my attention to errors or omissions in any part of this paper.[6]

Notation

The following notation is employed in Table I; also in Table II, except as specifically restricted.

a, b, c = positive reals.

br x = branch x. For each multiple-valued function, branches are designated in one or more different ways. When no branch designation is given, branch zero is to be understood.

$C(z)$ = $\int_0^z \cos(\frac{1}{2}\pi z^2)dz = -C(-z)$. $C(\pm\infty) = \pm\frac{1}{2}$.

$\text{cis}(z)$ = $\cos z + i \sin z = \exp(iz) = e^{iz}$ = cisoidal oscillation if $z = 2\pi ft$.

$D_\nu(z)$ = parabolic cylinder function of order ν.
$D_n(z) = \exp(-\frac{1}{4}z^2)H_n(z)$. $D_{-\frac{1}{2}}(z) = (2\pi)^{-\frac{1}{2}}z^{\frac{1}{2}}K_{\frac{1}{4}}(\frac{1}{4}z^2)$.
$D_{-1}(z) = (\frac{1}{2}\pi)^{\frac{1}{2}}\exp(\frac{1}{4}z^2)\,\text{erfc}(2^{-\frac{1}{2}}z)$.

$\text{erf}(z)$ = $\frac{2}{\sqrt{\pi}}\int_0^z \exp(-z^2)dz = -\text{erf}(-z)$. $\text{erf}(\pm\infty) = \pm 1$.

$\text{erfc}(z)$ = $\frac{2}{\sqrt{\pi}}\int_z^\infty \exp(-z^2)dz = 1 - \text{erf}(z)$.

f = frequency; parameter for the cisoidal oscillation. $-\infty < f < \infty$.

$F(f)$ = coefficient for cisoidal oscillation, parameter f.

$F_n(f)$ = coefficient of an i^n-multiple pair $(F_n(f), i^nF_n(g))$ in pairs (223)–(225).

[6] I am already much indebted to M. Paul Lévy for a number of suggestions including the expression of the general identical pair as the sum of any pair having even coefficients and its transposed pair.

g = epoch; parameter for the unit impulse. $-\infty < g < \infty$.

$g_0 < g < g_1$ restricts the given coefficient to the indicated limits; outside these limits the coefficient is zero.

$G(g)$ = coefficient for unit impulse, parameter g.

$H_n(z) = z^n - \frac{n(n-1)}{2} z^{n-2} + \frac{n(n-1)(n-2)(n-3)}{2 \cdot 4} z^{n-4} - \cdots$

= Hermite polynomial of order n.

$H_\nu^{(1)}(z) = \frac{2}{\pi} i^{-\nu-1} K_\nu(-iz)$ = Bessel function of the third kind.

$H_\nu^{(2)}(z) = \frac{2}{\pi} i^{\nu+1} K_\nu(iz)$ = Bessel function of the third kind.

$I(z)$ = imaginary part of z. $z = R(z) + iI(z)$.

$I_\nu(z) = i^{-\nu} J_\nu(iz)$.

j, k, l = integers greater than zero.

$J_\nu(z) = i^\nu I_\nu(-iz)$ = Bessel function of the first kind.

$K_\nu(z) = \frac{1}{2}\pi i^{\nu+1} H_\nu^{(1)}(iz)$.

m, n = positive integers, including zero.

$\mathcal{M}(\ )$ = mate of ( ).

p = $i2\pi f$, the imaginary radian frequency.

r, s = reals, positive or zero.

$R(z)$ = real part of z. $z = R(z) + iI(z)$.

$S(z) = \int_0^z \sin(\frac{1}{2}\pi z^2) dz = -S(-z)$. $S(\pm\infty) = \pm\frac{1}{2}$.

$\mathbb{S}_v(x) = \lim_{a \to \infty} aD_x^v \exp(-\pi a^2 x^2)$ = vth singularity function.

$\mathbb{S}_{-n}(x) = \left(\lambda_1 \pm \frac{1}{2(n-1)!}\right) x^{n-1} + \lambda_2 x^{n-2} + \cdots + \lambda_n$, $0 < \pm x$, $0 < n$.

t = time. $-\infty < t < \infty$.

v, w = integers, positive, negative or zero.

x, y = reals, unrestricted.

Y = admittance of system for cisoidal oscillation.

$Y_\nu(z) = \frac{1}{2}i[H_\nu^{(2)}(z) - H_\nu^{(1)}(z)]$ = Bessel function of the second kind.

z = complex quantity, unrestricted.

$\bar{z}$ = conjugate of z.

z^{μ}, br x = $\exp[\mu R \ (\log z) + i\mu \arg z]$, where $(2x - 1)\pi < \arg z \leqq (2x + 1)\pi$.

= $e^{i2\pi\mu v} z^{\mu}$, br$(x - v)$. Branches $(x + v)$, $v = 0, \pm 1, \pm 2, \cdots$ form a complete set and without repetition unless μ is a rational real.

$\alpha, \beta, \gamma, \delta$ = complex quantities, real parts greater than zero.

θ = principal argument. $-\pi < \theta \leqq \pi$.

λ, μ, ν = complex quantities, unrestricted.

ρ, σ, τ = complex quantities, real parts not less than zero.

$\phi_n(x)$ = $\exp(\pi x^2) D_x{}^n \exp(-2\pi x^2)$

= $(-2\pi^{\frac{1}{2}})^n D_n(2\pi^{\frac{1}{2}} x)$ where D_n is the parabolic cylinder function of order n

= $(-2\pi^{\frac{1}{2}})^n \exp(-\pi x^2) H_n(2\pi^{\frac{1}{2}} x)$ where H_n is the Hermite polynomial of order n.

$\psi(z)$ = $\Gamma'(z)/\Gamma(z)$ = logarithmic derivate of the gamma function. $-\psi(1)$ = Euler's constant = $0.5772\cdots$.

* marks a pair as being the limit approached by regular pairs.

	Not Restricted	Real Part	
		$\geqq 0$	> 0
Integers	v, w	m, n	j, k, l
Reals	f, g, t, x, y	r, s	a, b, c
Complex	z, λ, μ, ν	ρ, σ, τ	$\alpha, \beta, \gamma, \delta$

TABLE I

PAIRED COEFFICIENTS FOR THE CISOIDAL OSCILLATION AND THE UNIT IMPULSE [1]

Part 1. General Processes for Deriving the Mate

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation $\operatorname{cis}(2\pi ft) = \exp(pt)$	Coefficient $G(g)$ for the Unit Impulse $\mathfrak{S}_0(t-g) = \lim_{a\to 0} \frac{1}{a} e^{-\pi(t-g)^2/a^2}$
101.	F	$\int_{-\infty}^{\infty} F(f)\operatorname{cis}(2\pi fg)df$
102.	$\int_{-\infty}^{\infty} G(g)\operatorname{cis}(-2\pi fg)dg$	G
103.	F	$D_g \int_{-\infty}^{\infty} F(f)\operatorname{cis}(2\pi fg)p^{-1}df$
104.*	$\lambda_1(p-p_0) + \lambda_2(p-p_0)^2 + \cdots$, lim by 401*	$\operatorname{cis}(2\pi f_0 g)[\lambda_1\mathfrak{S}_1(g) + \lambda_2\mathfrak{S}_2(g) + \cdots]$
105.*	$\lambda_1\frac{1}{(p-p_0)} + \lambda_2\frac{1}{(p-p_0)^2} + \cdots$, lim by 408*	$\operatorname{cis}(2\pi f_0 g)\left(\lambda_1 + \lambda_2\frac{g}{1!} + \lambda_3\frac{g^2}{2!} + \lambda_4\frac{g^3}{3!} + \cdots\right)$, $0 < g$
106.*	$\lambda_1 p + \lambda_2 p^2 + \lambda_3 p^3 + \cdots$, lim by 401*	$\lambda_1\mathfrak{S}_1(g) + \lambda_2\mathfrak{S}_2(g) + \lambda_3\mathfrak{S}_3(g) + \cdots$
107.*	$\lambda_1\frac{1}{p} + \lambda_2\frac{1}{p^2} + \lambda_3\frac{1}{p^3} + \cdots$, lim by 408*	$\lambda_1 + \lambda_2\frac{g}{1!} + \lambda_3\frac{g^2}{2!} + \lambda_4\frac{g^3}{3!} + \cdots$, $0 < g$
108.*	$\frac{1}{p^a}\left(\lambda_0 + \lambda_1\frac{1}{p} + \lambda_2\frac{1}{p^2} + \cdots\right)$, $0 < a < 1$, lim by 516*	$\frac{1}{\Gamma(a)g^{1-a}}\left[\lambda_0 + \lambda_1\frac{1}{a}g + \lambda_2\frac{1}{a(a+1)}g^2 + \cdots\right]$, $0 < g$

[1] The pair for the oscillation cis $(-2\pi ft)$ and the unit impulse is $F(f)$, $G(-g)$ which differs from the tabulated pair only in the sign of the epoch; similarly, for the oscillation cos $(2\pi ft)$ or sin $(2\pi ft)$ the only change is the substitution for $G(g)$ of the even part of $G(g)$ or of the odd part of $-iG(g)$, the pairs being $F(f)$, $\frac{1}{2}[G(g) + G(-g)]$, or $F(f)$, $-i\frac{1}{2}[G(g) - G(-g)]$, respectively. Every pair in Table I may be thrown into the form of an evaluated Fourier integral by equating the pair after writing either $\int_{-\infty}^{\infty} df$ cis $(2\pi fg)$ before the coefficient $F(f)$ or $\int_{-\infty}^{\infty} dg$ cis $(-2\pi fg)$ before the coefficient $G(g)$. Every pair in Table I may also be regarded as an operational expression $F(p/i2\pi)$ of the operator $p = i2\pi f = d/dg$ with $G(g)$ its explicit expression in g.

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
109.*	$\frac{1}{p^a}(\lambda_0 + \lambda_1 p + \lambda_2 p^2 + \cdots)$, $0 < a < 1$, lim by 501*	$\frac{1}{\Gamma(a)g^{1-a}}\left[\lambda_0 - \lambda_1(1-a)\frac{1}{g} + \lambda_2(1-a)(2-a)\frac{1}{g^2} + \cdots\right]$, $0 <$ g
110.*	$\frac{1}{\sqrt{p}}\left(\lambda_0 + \lambda_1\frac{1}{p} + \lambda_2\frac{1}{p^2} + \lambda_3\frac{1}{p^3} + \cdots\right)$, lim by 518*	$\frac{1}{\sqrt{\pi g}}\left[\lambda_0 + \lambda_1\frac{2g}{1} + \lambda_2\frac{(2g)^2}{1\cdot 3} + \lambda_3\frac{(2g)^3}{1\cdot 3\cdot 5} + \cdots\right]$, $0 <$ g
111.*	$\frac{1}{\sqrt{p}}(\lambda_0 + \lambda_1 p + \lambda_2 p^2 + \lambda_3 p^3 + \cdots)$, lim by 502*	$\frac{1}{\sqrt{\pi g}}\left[\lambda_0 - \lambda_1\frac{1}{2g} + \lambda_2\frac{1\cdot 3}{(2g)^2} - \lambda_3\frac{1\cdot 3\cdot 5}{(2g)^3} + \cdots\right]$, $0 <$ g
112.	$\lambda_0\phi_0(f) + \lambda_1\phi_1(f) + \lambda_2\phi_2(f) + \cdots$	$\lambda_0\phi_0(g) + i\lambda_1\phi_1(g) + i^2\lambda_2\phi_2(g) + \cdots$

Part 2. Elementary Combinations and Transformations

201.	$F_1 \pm F_2$	$G_1 \pm G_2$
202.[2]	F_1F_2	$\int_{-\infty}^{\infty} G_1(x)G_2(g-x)dx$
203.	$\int_{-\infty}^{\infty} F_1(-x)F_2(f+x)dx$	G_1G_2
204.	λF	λG

[2] From (202) or (203), with g (or f) $= 0$, and (215) and (217) follow the important identities for th integrated product of two pairs of coefficients and for the integrated squared moduli of a pair of coefficients

$$\int_{-\infty}^{\infty} F_1(f)F_2(\pm f)df = \int_{-\infty}^{\infty} G_1(g)G_2(\mp g)dg,$$

$$\int_{-\infty}^{\infty} |F^2|df = \int_{-\infty}^{\infty} |G^2|dg,$$

$$\int_{-\infty}^{\infty} F_1(x)G_2(x)dx = \int_{-\infty}^{\infty} G_1(x)F_2(x)dx.$$

The symmetry of these identities is to be noted; this would not be the case if the radian frequenc $2\pi f$ were employed in place of the cyclic frequency f.

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
105.	$F(af)$	$\frac{1}{a}G\left(\frac{g}{a}\right)$
106.	$F(f - f_0) = F\left(\frac{p - p_0}{i2\pi}\right)$	$\text{cis}(2\pi f_0 g)G = e^{p_0 g}G$
107.	$\text{cis}(-2\pi f g_0)F = e^{-pg_0}F$	$G(g - g_0)$
108.	pF	$D_g G$
109.	$D_p F = \frac{1}{i2\pi}D_f F$	$-gG$
110.	$\frac{1}{p}F$	$\int_{-\infty}^{g} G dg = D_g^{-1}G$
111.	$\int_{-i\infty}^{p} F dp = i2\pi\int_{-\infty}^{f} F df = D_p^{-1}F$	$-\frac{1}{g}G$
112.	$D_\lambda F$	$D_\lambda G$
113.	$\int_{\lambda_0}^{\lambda} F d\lambda = D_\lambda^{-1}F$	$\int_{\lambda_0}^{\lambda} G d\lambda = D_\lambda^{-1}G$
114.	$F(-f)$	$G(-g)$
115.	$\bar{F}(\pm f)$	$\bar{G}(\mp g)$
116.	$F(f) \pm \bar{F}(f)$	$G(g) \pm \bar{G}(-g)$
117.	$G(\pm f)$	$F(\mp g)$
118.	$G(\pm ip)$	$\frac{1}{2\pi}F\left(\frac{\pm g}{2\pi}\right)$
119.	$\frac{F(f)}{p - p_0}$	$e^{p_0 g}\int_{-\infty}^{g} e^{-p_0 g}G(g)dg$
120.	$\frac{p}{p - p_0}F(f)$	$G(g) + p_0 e^{p_0 g}\int_{-\infty}^{g} e^{-p_0 g}G(g)dg$

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
221.	$f^{\pm\frac{1}{2}}D_f^{\pm 1}(f^{\pm\frac{1}{2}}F) = (\frac{1}{2} + fD_f)^{\pm 1}F$	$-g^{\pm\frac{1}{2}}D_g^{\pm 1}(g^{\pm\frac{1}{2}}G) = -(\frac{1}{2} + gD_g)^{\pm 1}G$
222.	$e^{\pm\pi f^2}D_f^{\nu}(e^{\mp\pi f^2}F) = (\mp 2\pi f + D_f)^{\nu}F$	$i^{\pm\nu}e^{\pm\pi g^2}D_g^{\nu}(e^{\mp\pi g^2}G) = i^{\pm\nu}(\mp 2\pi g + D_g)^{\nu}G$
223.[3]	$F_n(f)$, $n = 0, 1, 2, \cdots, 11$ where $F_n(f) = \frac{1}{4}[F(f) + i^{2n}F(-f) + i^{-n}G(f) + i^nG(-f)]$, $F_{n+4} = R(F_n)$, $F_{n+8} = I(F_n)$, $n = 0, 1, 2, 3$	$i^nF_n(g)$
224.	$F_0(f) + F_1(f) + F_2(f) + F_3(f)$ where F_n is as for 223.	$F_0(g) + iF_1(g) - F_2(g) - iF_3(g)$
225.	$F_4(f) + F_6(f) + F_5(f) + F_7(f)$ $+ i[F_8(f) + F_{10}(f) + F_9(f) + F_{11}(f)]$ where F_n is as for 223.	$F_4(g) - F_6(g) - F_9(g) + F_{11}(g)$ $+ i[F_8(g) - F_{10}(g) + F_5(g) - F_7(g)]$.
	Part 3. Key Pairs	
301.	$\sec p$; $[\gamma(p + \lambda)$ for p, if $-\frac{1}{2}\pi R(1/\gamma) < R(\lambda) < \frac{1}{2}\pi R(1/\gamma)]$	$\frac{1}{2}\operatorname{sech}(\frac{1}{2}\pi g)$
302.	$p^{\frac{1}{2}(\alpha-1)}K_{\alpha-1}(\sqrt{p})$; $[\gamma(p + \rho)$ for p, and ν for α with $(p + \beta)$ for $p]$	$(2g)^{-\alpha}\exp\left(-\frac{1}{4g}\right)$, $0 <$
303.	$\frac{1}{p}\exp\left(-\frac{a^2+1}{p}\right)I_{\alpha-1}\left(\frac{2a}{p}\right)$; $[\gamma(p + \beta)$ for p, and $\sqrt{\gamma\delta}$ for a with $(p + \beta)$ for $p]$	$J_{\alpha-1}(2a\sqrt{g})J_{\alpha-1}(2\sqrt{g})$, $0 <$
304.	$\exp(\frac{1}{4}p^2)D_{-\alpha}(p)$; $[\sqrt{\gamma}(p + \rho)$ for $p]$	$\frac{1}{\Gamma(\alpha)}g^{\alpha-1}\exp(-\frac{1}{2}g^2)$, $0 <$

The coefficients of the i^n-multiple pairs satisfy the following integral equations:

$$F_n(f) = (-1)^{\frac{1}{2}}\ 2\int_0^\infty F_n(g)\cos(2\pi fg)dg, \quad n = 0, 2, 4, 6, 8, 10$$

$$F_n(f) = (-1)^{\frac{1}{2}(n-1)}2\int_0^\infty F_n(g)\sin(2\pi fg)dg, \quad n = 1, 3, 5, 7, 9, 11$$

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
305.	$p^{-\frac{1}{2}\alpha}\exp\left(\frac{1}{4p}\right)D_{-\alpha}\left(\frac{1}{\sqrt{p}}\right)$; $[\gamma(p+\rho)$ for $p]$	$\frac{1}{\Gamma(\alpha)}(2g)^{\frac{1}{2}\alpha-1}\exp(-\sqrt{2g}),$ $0<g$
306.	$(p^2+b^2)^{\frac{1}{4}-\frac{1}{2}\alpha}K_{\alpha-\frac{1}{2}}(\sqrt{p^2+b^2})$, $0<R(\alpha)<\frac{3}{2}$; [$(p+\rho)$ for p, and δ for b without the restriction on α with $(p+\beta)$ for p, if $-R(\beta)<R(i\delta)<R(\beta)$]	$\sqrt{\frac{\pi}{2}}\left(\frac{g^2-1}{b^2}\right)^{\frac{1}{2}\alpha-\frac{1}{4}}J_{\alpha-1}(b\sqrt{g^2-1}),$ $1<g$
307.	$(\delta^2-p^2)^{\frac{1}{4}-\frac{1}{2}\alpha}J_{\alpha-\frac{1}{2}}(\sqrt{\delta^2-p^2})$; $[(p+\lambda)$ for $p]$	$\frac{1}{\sqrt{2\pi}}\left(\frac{1-g^2}{\delta^2}\right)^{\frac{1}{2}\alpha-\frac{1}{4}}J_{\alpha-1}(\delta\sqrt{1-g^2}),$ $-1<g<1$

Part 4. Rational Algebraic Functions of f.

401.*	$p^n=\lim_{a\to 0}p^ne^{-\pi a^2f^2}$	$\mathbb{S}_n(g)=\lim_{a\to 0}\frac{1}{a}D_g{}^ne^{-\pi a^{-2}g^2}$
402.*	$p^n=\lim_{a\to 0}\sum_{k=1}^{n+1}\frac{(-1)^{k-1}(n+1)!(1-e^{-kap})}{k!(n-k+1)!\,a^{n+1}p}$	$\mathbb{S}_n(g)$
403.*	$1=\lim_{\beta\to 0}e^{-\beta\lvert p-p_0\rvert}$	$\mathbb{S}_0(g)$, unit impulse at $g=0$
404.*	$p=\lim_{\beta\to 0}pe^{-\pi\beta f^2}$	$\mathbb{S}_1(g)$, negative unit doublet at $g=0$
405.*	$\lvert p^{2n}\rvert=\lim_{\beta\to 0}D_\beta{}^{2n}e^{-\beta\lvert p\rvert}$	$(-1)^n\mathbb{S}_{2n}(g)$
406.*	$\lvert p^{2n+1}\rvert=-\lim_{\beta\to 0}D_\beta{}^{2n+1}e^{-\beta\lvert p\rvert}$	$\frac{(-1)^{n+1}(2n+1)!}{\pi g^{2n+2}}$
407.*	$\lvert p\rvert$, lim by 406*	$-\frac{1}{\pi g^2}$
408.*	$p^{-n}=\lim_{\beta\to 0}(p+\beta)^{-n}$, $0<n$	$\frac{g^{n-1}}{(n-1)!}$, $0<g$
409.*	$p^{-n}=\lim_{\beta\to 0}(p-\beta)^{-n}$, $0<n$	$\frac{-g^{n-1}}{(n-1)!}$, $g<0$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
410.*	$p^{-n} = \lim_{a\to 0} p^{-n}e^{-\pi a^2 f^2}, \quad 0 < n$	$\mathfrak{S}_{-n}(g) = \lim_{a\to 0} \frac{1}{a} D_g{}^{-n} e^{-\pi a^{-2} g^2}$ $= \left(\lambda_1 \pm \frac{1}{2(n-1)!}\right) g^{n-1} + \lambda_2 g^{n-2} + \lambda_3 g^{n-3} + \cdots + \lambda_n, \quad 0 < \pm g$
411.*	$p^{-n} = \lim_{\beta\to 0}\left[\frac{\frac{1}{2}}{(p-\beta)^n} + \frac{\frac{1}{2}}{(p+\beta)^n} + \sum_{k=1}^{n}\left(\frac{\lambda_k(n-k)!}{(p+\beta)^{n-k+1}} - \frac{\lambda_k(n-k)!}{(p-\beta)^{n-k+1}}\right)\right], \quad 0 < n$	$\mathfrak{S}_{-n}(g)$
415.*	$\frac{1}{p} = \lim_{\beta\to 0}\left(\frac{\frac{1}{2}-\lambda}{p-\beta} + \frac{\frac{1}{2}+\lambda}{p+\beta}\right)$	$\mathfrak{S}_{-1}(g) = \lambda \pm \frac{1}{2}$, $0 < \pm g$, unit step at $g = 0$
416.*	$\frac{1}{p^2} = \lim_{\beta\to 0}\left[\frac{\frac{1}{2}-\lambda}{(p-\beta)^2} + \frac{\frac{1}{2}+\lambda}{(p+\beta)^2} - \frac{2\beta\mu}{p^2-\beta^2}\right]$	$\mathfrak{S}_{-2}(g) = \frac{1}{2}\lvert g\rvert + \lambda g + \mu$
421.*	$F(f) = \lim_{a\to 0} F_1(f)$, where F is any proper rational fraction in p with n distinct poles, the degree of pole z_j being n_j. All pure imaginary poles in F are the limits of corresponding poles in F_1 which have assigned real parts $\pm a$.	$\sum_{j=1}^{n}\sum_{k=1}^{n} \pm \lambda_{jk} e^{z_j g} g^{n_j-k}$, $0 < \pm g, \pm R(z_j) < 0$, where $\lambda_{jk} = \frac{\{D_p{}^{k-1}[F(f)(p-z_j)^{n_j}]\}_{p=z}}{(k-1)!(n_j-k)!}$ and the upper or lower signs for each term are employed according as the real part of z_j (either actual or vestigial) is less than or greater than zero.
431.	$\frac{1}{(p+\beta)^n}, \quad 0 < n$	$\frac{1}{(n-1)!} e^{-\beta g} g^{n-1}, \quad 0 < g$
432.	$\frac{1}{(p-\beta)^n}, \quad 0 < n$	$\frac{-1}{(n-1)!} e^{\beta g} g^{n-1}, \quad g < 0$
433.	$\frac{1}{(p^2-\beta^2)^n}, \quad 0 < n$	$\frac{(-1)^n}{(n-1)!} e^{-\beta\lvert g\rvert} \sum_{k=1}^{n} \frac{(n+k-2)!\lvert g^{n-k}\rvert}{(k-1)!(n-k)!(2\beta)^{n+k-1}}$ $= \frac{(-1)^n \lvert g^{n-\frac{1}{2}}\rvert K_{n-\frac{1}{2}}(\beta\lvert g\rvert)}{(n-1)!\,\pi^{\frac{1}{2}}(2\beta)^{n-\frac{1}{2}}}$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation		Coefficient $G(g)$ for the Unit Impulse	
38.	$\frac{1}{p+\beta}$		$e^{-\beta g}$,	$0 < g$
39.	$\frac{1}{p-\beta}$		$-e^{\beta g}$,	$g < 0$
40.*	$\frac{1}{p-p_0}$,	lim by 415*	$\operatorname{cis}(2\pi f_0 g)\mathfrak{S}_{-1}(g)$	
41.*	$\frac{p}{p+\beta}$,	lim by 403*	$\mathfrak{S}_0(g) - \beta e^{-\beta g}$,	$0 < g$
42.	$\frac{1}{(p+\beta)^2}$		$ge^{-\beta g}$,	$0 < g$
43.	$\frac{1}{(p-\beta)^2}$		$-ge^{\beta g}$,	$g < 0$
44.	$\frac{1}{p^2-\beta^2}$		$-\frac{1}{2\beta}e^{-\beta\lvert g\rvert}$	
45.	$\frac{p}{p^2-\beta^2}$		$\pm\frac{1}{2}e^{\mp\beta g}$,	$0 < \pm g$
46.*	$\frac{a}{p^2+a^2}$,	lim by 415*	$\sin ag\ \mathfrak{S}_{-1}(g)$	
47.*	$\frac{p}{p^2+a^2}$,	lim by 415*	$\cos ag\ \mathfrak{S}_{-1}(g)$	
48.	$\frac{1}{(p+\alpha)(p+\beta)}$		$\frac{e^{-\beta g}-e^{-\alpha g}}{\alpha-\beta}$,	$0 < g$
49.	$\frac{p}{(p+\alpha)(p+\beta)}$		$\frac{\alpha e^{-\alpha g}-\beta e^{-\beta g}}{\alpha-\beta}$,	$0 < g$
50.	$\frac{1}{(p+\beta)^3}$		$\frac{1}{2}g^2e^{-\beta g}$,	$0 < g$
51.	$\frac{1}{(p-\beta)^3}$		$-\frac{1}{2}g^2e^{\beta g}$,	$g < 0$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse	
452.	$\dfrac{1}{(p+\alpha)(p+\beta)(p+\gamma)}$	$\dfrac{(\gamma-\beta)e^{-\alpha g}+(\alpha-\gamma)e^{-\beta g}+(\beta-\alpha)e^{-\gamma g}}{(\alpha-\beta)(\beta-\gamma)(\gamma-\alpha)}$,	$0 <$
453.	$\dfrac{p}{(p+\alpha)(p+\beta)(p+\gamma)}$	$\dfrac{\alpha(\beta-\gamma)e^{-\alpha g}+\beta(\gamma-\alpha)e^{-\beta g}+\gamma(\alpha-\beta)e^{-\gamma g}}{(\alpha-\beta)(\beta-\gamma)(\gamma-\alpha)}$	$0 <$
454.	$\dfrac{\alpha\beta}{p(p+\alpha)(p+\beta)}-\dfrac{1}{p}$	$\dfrac{\beta e^{-\alpha g}-\alpha e^{-\beta g}}{\alpha-\beta}$,	$0 <$
455.*	$\dfrac{a^2}{p(p^2+a^2)}$,	lim by 415* $2\sin^2(\frac{1}{2}ag)\mathfrak{S}_{-1}(g)$	
456.	$\dfrac{1}{(p+b+ia)(p+b-ia)}$	$\dfrac{1}{a}\sin ag\, e^{-bg}$,	$0 <$
457.	$\dfrac{1}{(p-b+ia)(p-b-ia)}$	$-\dfrac{1}{a}\sin ag\, e^{bg}$,	$g <$
458.	$\dfrac{p}{(p+b+ia)(p+b-ia)}$	$\left(\cos ag-\dfrac{b}{a}\sin ag\right)e^{-bg}$,	$0 <$
459.	$\dfrac{p}{(p-b+ia)(p-b-ia)}$	$-\left(\cos ag+\dfrac{b}{a}\sin ag\right)e^{bg}$,	$g <$

Part 5. Irrational Algebraic Functions of f.

501.*	$p^{n-\alpha}=\lim_{\beta\to 0}p^{n-\alpha}e^{-\beta\lvert p\rvert}$,	$0 < R(\alpha) < 1$	$\dfrac{1}{\Gamma(\alpha-n)}g^{\alpha-n-1}$,	$0 <$
502.*	$p^{n-\frac{1}{2}}=\lim_{\beta\to 0}p^{n-\frac{1}{2}}e^{-\beta\lvert p\rvert}$,	$0 < n$	$\dfrac{(-1)^n 1\cdot 3\cdot 5\cdots(2n-1)}{(\frac{1}{2}\pi)^{\frac{1}{2}}}(2g)^{-n-\frac{1}{2}}$,	$0 <$
503.*	$p^{\frac{1}{2}}$,	lim by 502*	$-\frac{1}{2}\pi^{-\frac{1}{2}}g^{-\frac{3}{2}}$,	$0 <$
504.*	$p^{\frac{3}{2}}$,	lim by 502*	$\frac{3}{4}\pi^{-\frac{1}{2}}g^{-\frac{5}{2}}$,	$0 <$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation		Coefficient $G(g)$ for the Unit Impulse	
05.*	$\lvert f^{\frac{1}{2}}\rvert$,	lim by 503*	$\frac{-1}{4\pi}\lvert g^{-\frac{3}{2}}\rvert$	
06.*	$(p+\beta)^{\frac{1}{2}}$,	lim by 820	$-\frac{1}{2}\pi^{-\frac{1}{2}}e^{-\beta g}g^{-\frac{3}{2}}$,	$0<g$
16.*	$p^{-\alpha}=\lim_{\beta\to 0}(p+\beta)^{-\alpha}$,	br 0	$\frac{1}{\Gamma(\alpha)}g^{\alpha-1}$,	br 0, $0<g$
17.*	$p^{-\alpha}=\lim_{\beta\to 0}(p-\beta)^{-\alpha}$,	br $(-\frac{1}{2})$	$\frac{-1}{\Gamma(\alpha)}g^{\alpha-1}$,	br 0, $g<0$
18.*	$p^{-n-\frac{1}{2}}=\lim_{\beta\to 0}(p+\beta)^{-n-\frac{1}{2}}$,	br 0, $0<n$	$\frac{(\frac{1}{2}\pi)^{-\frac{1}{2}}}{1\cdot 3\cdot 5\cdots(2n-1)}(2g)^{n-\frac{1}{2}}$,	br 0, $0<g$
19.*	$p^{-n-\frac{1}{2}}=\lim_{\beta\to 0}(p-\beta)^{-n-\frac{1}{2}}$,	br $\frac{1}{2}$, $0<n$	$\frac{(\frac{1}{2}\pi)^{-\frac{1}{2}}}{1\cdot 3\cdot 5\cdots(2n-1)}(2g)^{n-\frac{1}{2}}$,	br 0, $g<0$
20.*	$p^{-\frac{3}{2}}$,	lim by 518*	$2(g/\pi)^{\frac{1}{2}}$,	$0<g$
21.	$p^{-\alpha}$,	$0<R(\alpha)<1$	$\frac{1}{\Gamma(\alpha)}g^{\alpha-1}$,	$0<g$
22.	$p^{-\frac{1}{2}}$		$(\pi g)^{-\frac{1}{2}}$,	$0<g$
23.	$\lvert f^{-\frac{1}{2}}\rvert$		$\lvert g^{-\frac{1}{2}}\rvert$	
24.	$(p+\beta)^{-\alpha}$,	br v	$\frac{1}{\Gamma(\alpha)}e^{-i2\pi\alpha(v+w)}e^{-\beta g}g^{\alpha-1}$,	br w, $0<g$
25.	$(p-\beta)^{-\alpha}$,	br $(v-\frac{1}{2})$	$\frac{-1}{\Gamma(\alpha)}e^{-i2\pi\alpha(v+w)}e^{\beta g}g^{\alpha-1}$,	br w, $g<0$
26.	$(p+\beta)^{-\frac{1}{2}}$,	br 0	$e^{-\beta g}(\pi g)^{-\frac{1}{2}}$,	br 0, $0<g$
27.	$(p-\beta)^{-\frac{1}{2}}$,	br $\frac{1}{2}$	$e^{\beta g}(\pi g)^{-\frac{1}{2}}$,	br 0, $g<0$
28.	$(p-\beta)^{-\frac{1}{2}}$,	br 0	$-e^{\beta g}(\pi g)^{-\frac{1}{2}}(\operatorname{erf}\sqrt{\beta g}-\frac{1}{2}\mp\frac{1}{2})$,	br 0, $0<\pm g$
29.	$(p+\beta)^{-\frac{3}{2}}$,	br 0	$2e^{-g}(g/\pi)^{\frac{1}{2}}$,	br 0, $0<g$
30.	$(p+\beta)^{1-\alpha}-(p+\gamma)^{1-\alpha}$		$\frac{1}{\Gamma(\alpha-1)}g^{\alpha-2}(e^{-\beta g}-e^{-\gamma g})$,	$0<g$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
541.	$\dfrac{\sqrt{p}}{p+\gamma}$	$\dfrac{1}{\sqrt{\pi g}} + \sqrt{-\gamma}\, e^{-\gamma g} \operatorname{erf} \sqrt{-\gamma g}, \quad 0 <$
542.	$\dfrac{1}{(p+\gamma)\sqrt{p}}$	$\dfrac{1}{\sqrt{-\gamma}} e^{-\gamma g} \operatorname{erf} \sqrt{-\gamma g}, \quad 0 <$
543.	$-\dfrac{1}{1+\sqrt{\beta p}}$	$\dfrac{1}{\sqrt{\pi\beta g}} - \dfrac{1}{\beta} \exp \dfrac{g}{\beta} \operatorname{erfc} \sqrt{\dfrac{g}{\beta}}, \quad 0 <$
544.*	$\dfrac{p}{1+\sqrt{\beta p}}$	$-\dfrac{1}{\beta} \mathfrak{S}_0(g) + \dfrac{2g-\beta}{2\beta g\sqrt{\pi\beta g}} - \dfrac{1}{\beta^2} \exp \dfrac{g}{\beta} \operatorname{erfc} \sqrt{\dfrac{g}{\beta}}, \quad 0 <$
545.	$\dfrac{p}{(p+\gamma)(1+\sqrt{\beta p})}$	$\dfrac{-\gamma e^{-\gamma g}}{1+\beta\gamma}(1 - \sqrt{-\beta\gamma} \operatorname{erf} \sqrt{-\gamma g}) + \dfrac{1}{\sqrt{\pi\beta g}} - \dfrac{\exp (g/\beta)}{\beta(1+\beta\gamma)} \operatorname{erfc} \sqrt{\dfrac{g}{\beta}}, \quad 0 <$
546.	$\dfrac{1}{(p+\gamma)\sqrt{p+\beta}}$	$\dfrac{1}{\sqrt{\beta-\gamma}} e^{-\gamma g} \operatorname{erf} \sqrt{(\beta-\gamma)g}, \quad 0 <$
547.	$\dfrac{\sqrt{\beta}}{p\sqrt{p+\beta}} - \dfrac{1}{p}$	$-\operatorname{erfc} \sqrt{\beta g}, \quad 0 <$
548.	$\dfrac{1}{p}\sqrt{\dfrac{p}{\beta}+1} - \dfrac{1}{p}$	$\dfrac{1}{\sqrt{\pi\beta g}} e^{-\beta g} - \operatorname{erfc} \sqrt{\beta g}, \quad 0 <$
549.	$\dfrac{\sqrt{p+\beta}}{p+\gamma}$	$\dfrac{1}{\sqrt{\pi g}} e^{-\beta g} + \sqrt{\beta-\gamma}\, e^{-\gamma g} \operatorname{erf} \sqrt{(\beta-\gamma)g}, \quad 0 <$
550.*	$\dfrac{\sqrt{p}}{1+\sqrt{\beta p}}$	$\dfrac{1}{\sqrt{\beta}} \mathfrak{S}_0(g) - \dfrac{1}{\beta\sqrt{\pi g}} + \dfrac{1}{\beta\sqrt{\beta}} \exp \dfrac{g}{\beta} \operatorname{erfc} \sqrt{\dfrac{g}{\beta}}, \quad 0 <$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
51.	$\dfrac{1}{\sqrt{p}(1+\sqrt{\beta p})}$	$\dfrac{1}{\sqrt{\beta}}\exp\dfrac{g}{\beta}\operatorname{erfc}\sqrt{\dfrac{g}{\beta}}$, $0<g$
52.	$\dfrac{\sqrt{p}}{(p+\gamma)(1+\sqrt{\beta p})}$	$\dfrac{\sqrt{-\gamma}\,e^{-\gamma g}}{1+\beta\gamma}(\operatorname{erf}\sqrt{-\gamma g}-\sqrt{-\beta\gamma})+\dfrac{\exp(g/\beta)}{\sqrt{\beta}(1+\beta\gamma)}\operatorname{erfc}\sqrt{\dfrac{g}{\beta}}$, $0<g$
53.*	$\sqrt{\dfrac{p+\alpha}{p}}$	$\mathbb{S}_0(g)+\dfrac{\alpha}{2}e^{-\frac{1}{2}\alpha g}[I_1(\frac{1}{2}\alpha g)+I_0(\frac{1}{2}\alpha g)]$, $0<g$
54.*	$\dfrac{1}{p}\sqrt{\dfrac{p+\alpha}{p}}$	$e^{-\frac{1}{2}\alpha g}[\alpha g I_1(\frac{1}{2}\alpha g)+(1+\alpha g)I_0(\frac{1}{2}\alpha g)]$, $0<g$
55.	$\dfrac{1}{\sqrt{(p+\alpha)(p+\beta)}}$	$e^{-\frac{1}{2}(\alpha+\beta)g}I_0[\frac{1}{2}(\alpha-\beta)g]$, $0<g$
56.*	$\sqrt{p^2+a^2}$	$\mathbb{S}_1(g)+\dfrac{a}{g}J_1(ag)$, $0<g$
57.	$\dfrac{1}{\sqrt{p^2+a^2}}$	$J_0(ag)$, $0<g$
58.	$\dfrac{1}{\sqrt{\beta^2-p^2}}$	$\dfrac{1}{\pi}K_0(\beta\lvert g\rvert)$
59.*	$\dfrac{\sqrt{p+\alpha}}{\sqrt{p}+\sqrt{p+\alpha}}$	$\frac{1}{2}\mathbb{S}_0(g)+\dfrac{1}{2g}e^{-\frac{1}{2}\alpha g}I_1(\frac{1}{2}\alpha g)$, $0<g$
60.	$\dfrac{\sqrt{p+\alpha}}{p(\sqrt{p}+\sqrt{p+\alpha})}-\dfrac{1}{p}$	$-\frac{1}{2}e^{-\frac{1}{2}\alpha g}[I_1(\frac{1}{2}\alpha g)+I_0(\frac{1}{2}\alpha g)]$, $0<g$
571.	$(p^2+b^2)^{-\alpha}$, $0<R(\alpha)<1$	$\dfrac{\sqrt{\pi}}{\Gamma(\alpha)}\left(\dfrac{g}{2b}\right)^{\alpha-\frac{1}{2}}J_{\alpha-\frac{1}{2}}(bg)$, $0<g$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
572.	$[(p+\beta)^2+r^2]^{-\alpha}$	$\frac{\sqrt{\pi}}{\Gamma(\alpha)}\left(\frac{g}{2r}\right)^{\alpha-\frac{1}{2}} e^{-\beta g} J_{\alpha-\frac{1}{2}}(rg), \quad 0 < g$
573.	$\sqrt{\frac{p}{p+\beta}}\,(\sqrt{p+\beta}+\sqrt{p})^{-2\alpha}$	$\frac{1}{4}\beta^{-\alpha+1} e^{-\frac{1}{2}\beta g}[I_{\alpha-1}(\frac{1}{2}\beta g) - 2I_\alpha(\frac{1}{2}\beta g) + I_{\alpha+1}(\frac{1}{2}\beta g)], \quad 0 < g$
574.	$\frac{(\sqrt{p+\beta}+\sqrt{p})^{-2\sigma}}{\sqrt{p(p+\beta)}}$	$\beta^{-\sigma} e^{-\frac{1}{2}\beta g} I_\sigma(\frac{1}{2}\beta g), \quad 0 < g$
575.	$\frac{[\sqrt{(p+\rho)^2+a^2}+(p+\rho)]^{-\alpha+1}}{\sqrt{(p+\rho)^2+a^2}}$	$a^{-\alpha+1} e^{-\rho g} J_{\alpha-1}(ag), \quad 0 < g$
576.	$[\sqrt{(p+\rho)^2+a^2}+(p+\rho)]^{-\alpha}$	$\frac{\alpha e^{-\rho g}}{a^\alpha g} J_\alpha(ag), \quad 0 < g$

Part 6. Exponential and Trigonometric Functions of f or f^{-1}.

601.*	e^{-rp}	$\mathfrak{S}_0(g-r)$
602.*	$\frac{e^{-rp}}{p}$	$\mathfrak{S}_{-1}(g-r)$
603.	$\frac{1}{p}(1-e^{-ap})$	$1, \quad 0 < g < a$
604.	$\frac{e^{-rp}}{p+\beta}$	$e^{-\beta(g-r)}, \quad r <$
611.	$\frac{\alpha}{\sin \alpha p} - \frac{1}{p}$	$\frac{1}{2}\tanh\frac{\pi g}{2\alpha} \mp \frac{1}{2}, \quad 0 < \pm$
612.*	$\tan \alpha p$	$\frac{-1}{2\alpha \sinh \frac{\pi g}{2\alpha}}$
613.*	$\alpha \operatorname{ctn} \alpha p - \frac{1}{p}$	$\frac{1}{2}\operatorname{ctnh}\frac{\pi g}{2\alpha} \mp \frac{1}{2}, \quad 0 < \pm$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation		Coefficient $G(g)$ for the Unit Impulse	
14.	$\dfrac{p}{\sin \alpha p}$		$\dfrac{\pi}{4\alpha^2 \cosh^2 \dfrac{\pi g}{2\alpha}}$	
15.	$\dfrac{\sin \beta p}{\sin \alpha p}$,	$R(\beta) < R(\alpha)$	$\dfrac{\dfrac{1}{2\alpha} \sin \dfrac{\pi\beta}{\alpha}}{\cosh \dfrac{\pi g}{\alpha} + \cos \dfrac{\pi\beta}{\alpha}}$	
16.	$\dfrac{\cos \beta p}{\cos \alpha p}$,	$R(\beta) < R(\alpha)$	$\dfrac{\dfrac{1}{\alpha} \cos \dfrac{\pi\beta}{2\alpha} \cosh \dfrac{\pi g}{2\alpha}}{\cosh \dfrac{\pi g}{\alpha} + \cos \dfrac{\pi\beta}{\alpha}}$	
17.	$\dfrac{\alpha \sin \beta p}{\beta p \sin \alpha p} - \dfrac{1}{p}$,	$R(\beta) < R(\alpha)$	$\dfrac{\alpha}{\pi\beta} \tan^{-1} \dfrac{\tanh \dfrac{\pi g}{2\alpha}}{\operatorname{ctn} \dfrac{\pi\beta}{2\alpha}} \mp \dfrac{1}{2}$,	$0 < \pm g$
18.	$\dfrac{\cos \beta p}{p \cos \alpha p} - \dfrac{1}{p}$,	$R(\beta) < R(\alpha)$	$-\dfrac{1}{\pi} \tan^{-1} \dfrac{\cos \dfrac{\pi\beta}{2\alpha}}{\sinh \dfrac{\pi g}{2\alpha}}$	
19.*	$\cosh(ap)$		$\frac{1}{2}[\mathbb{S}_0(g + a) + \mathbb{S}_0(g - a)]$	
20.	$\dfrac{\cosh(ap)}{p} - \dfrac{1}{p}$		$\mp \frac{1}{2}$,	$0 < \pm g < a$
21.	$\dfrac{\cosh(ap)}{(p - p_0)\cosh(ap_0)} - \dfrac{1}{p - p_0}$		$\frac{1}{2} \operatorname{cis}(2\pi f_0 g)[\tanh(ap_0) \mp 1]$,	$0 < \pm g < a$
22.	$\dfrac{\sinh(ap)}{p}$		$\frac{1}{2}$,	$-a < g < a$
23.	$\dfrac{\sinh(ap)}{ap^2} - \dfrac{1}{p}$		$\dfrac{g}{2a} \mp \frac{1}{2}$,	$0 < \pm g < a$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
624.	$\frac{p_0 \sinh(ap)}{p(p-p_0)\sinh(ap_0)} - \frac{1}{p-p_0}$	$\frac{1}{2}\{\mathrm{cis}(2\pi f_0 g)[\mathrm{ctnh}(ap_0) \mp 1] - \mathrm{csch}(ap_0)\}$ $0 < \pm g <$
625.	$\mathrm{sech}\ \pi f$	$\mathrm{sech}\ \pi g$
631.	$p^{\alpha-1}e^{-\beta\lvert p\rvert}$	$\frac{\Gamma(\alpha)}{i2\pi}[(-g-i\beta)^{-\alpha} - (-g+i\beta)^{-\alpha}]$
632.	$e^{-\beta\lvert p\rvert}$	$\frac{1}{\pi}\cdot\frac{\beta}{\beta^2+g^2}$
633.	$\frac{1}{p}e^{-\beta\lvert p\rvert} - \frac{1}{p}$	$-\frac{1}{\pi}\tan^{-1}\frac{\beta}{g}$
634.	$\frac{e^{-\beta\lvert p\rvert}}{\sqrt{\lvert p\rvert}}$	$\sqrt{\frac{\beta+\sqrt{\beta^2+g^2}}{2\pi(\beta^2+g^2)}}$
635.	$\lvert p\rvert e^{-\beta\lvert p\rvert}$	$\frac{\beta^2-g^2}{\pi(\beta^2+g^2)^2}$
645.*	$\sin(a\lvert p\rvert)$	$\frac{a}{\pi(a^2-g^2)}$
651.	$\frac{1}{\sqrt{p+\beta}}\exp\frac{\rho}{p+\beta}$	$\frac{e^{-\beta\rho}}{\sqrt{\pi g}}\cosh(2\sqrt{\rho g}),$ $0 <$
652.*	$\frac{1}{\sqrt{p}}\exp\frac{1}{4p}$	$\frac{1}{\sqrt{\pi g}}\cosh\sqrt{g},$ $0 <$
653.	$(p+\beta)^{-\frac{3}{2}}\exp\frac{\rho}{p+\beta}$	$\frac{e^{-\beta\rho}}{\sqrt{\pi\rho}}\sinh(2\sqrt{\rho g}),$ $0 <$
654.*	$\exp\left(-\frac{1}{p}\right)$	$\mathbb{S}_0(g) - \frac{1}{\sqrt{g}}J_1(2\sqrt{g}),$ $0 <$
655.	$\frac{1}{p}\exp\left(-\frac{1}{p}\right)$	$J_0(2\sqrt{g}),$ $0 <$
656.*	$\frac{1}{p^2}\exp\left(-\frac{1}{p}\right)$	$\sqrt{g}\,J_1(2\sqrt{g}),$ $0 <$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Part 7. Exponential and Trigonometric Functions of f^2.

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
701.	$e^{\rho p^2}, \qquad 0 < \lvert\rho\rvert$	$\frac{1}{2\sqrt{\pi\rho}}\, e^{-\frac{1}{4}g^2/\rho}$
702.	$\phi_n(f) = e^{\pi f^2} D_f{}^n e^{-2\pi f^2}$	$i^n \phi_n(g)$
703.	$\phi_0(f) = e^{-\frac{1}{4}x^2} = e^{-\pi f^2}, \quad x = f\sqrt{4\pi}$ in 703–711	$\phi_0(g) = e^{-\pi g^2}$
704.	$\phi_1(f) = -e^{-\frac{1}{4}x^2}(4\pi)^{\frac{1}{2}} x$	$i\phi_1(g)$
705.	$\phi_2(f) = e^{-\frac{1}{4}x^2}(4\pi)(x^2 - 1)$	$-\phi_2(g)$
706.	$\phi_3(f) = -e^{-\frac{1}{4}x^2}(4\pi)^{\frac{3}{2}}(x^3 - 3x)$	$-i\phi_3(g)$
707.	$\phi_4(f) = e^{-\frac{1}{4}x^2}(4\pi)^2(x^4 - 6x^2 + 3)$	$\phi_4(g)$
708.	$\phi_5(f) = -e^{-\frac{1}{4}x^2}(4\pi)^{\frac{5}{2}}(x^5 - 10x^3 + 15x)$	$i\phi_5(g)$
709.	$\phi_6(f) = e^{-\frac{1}{4}x^2}(4\pi)^3(x^6 - 15x^4 + 45x^2 - 15)$	$-\phi_6(g)$
710.	$\phi_7(f) = -e^{-\frac{1}{4}x^2}(4\pi)^{\frac{7}{2}}(x^7 - 21x^5 + 105x^3 - 105x)$	$-i\phi_7(g)$
711.	$\phi_8(f) = e^{-\frac{1}{4}x^2}(4\pi)^4(x^8 - 28x^6 + 210x^4 - 420x^2 + 105)$	$\phi_8(g)$
712.	$e^{-\pi f^2}(4\pi f^2 - 3)^2$	$e^{-\pi g^2}(4\pi g^2 - 3)^2$
713.	$p^n e^{-\pi\beta f^2}$	$\frac{1}{\sqrt{\beta}} D_g{}^n e^{-\pi g^2/\beta} = \frac{1}{\sqrt{\beta}(\sqrt{2\beta})^n} \times e^{-\frac{1}{2}\pi g^2/\beta} \phi_n\left(\frac{g}{\sqrt{2\beta}}\right)$
714.	$e^{-\pi\beta f^2}$	$\frac{1}{\sqrt{\beta}}\, e^{-\pi g^2/\beta}$
715.	$pe^{-\pi\beta f^2}$	$-\frac{2\pi g}{\beta\sqrt{\beta}}\, e^{-\pi g^2/\beta}$
716.	$p^2 e^{-\pi\beta f^2}$	$\frac{2\pi}{\beta^2\sqrt{\beta}}\, e^{-\pi g^2/\beta}(2\pi g^2 - \beta)$

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
717.	$p^3 e^{-\pi\beta f^2}$	$-\frac{4\pi^2 g}{\beta^3\sqrt{\beta}} e^{-\pi g^2/\beta}(2\pi g^2 - 3\beta)$
718.	$p^4 e^{-\pi\beta f^2}$	$\frac{4\pi^2}{\beta^4\sqrt{\beta}} e^{-\pi g^2/\beta}(4\pi^2 g^4 - 12\pi\beta g^2 + 3\beta^2)$
719.	$p^n e^{-\frac{1}{2}\pi f^2}$	$\sqrt{2} D_g{}^n e^{-2\pi g^2} = \sqrt{2} e^{-\pi g^2}\phi_n(g)$
720.	$e^{-\frac{1}{2}\pi f^2}$	$\sqrt{2} e^{-2\pi g^2}$
721.	$pe^{-\frac{1}{2}\pi f^2}$	$-4\pi g\sqrt{2} e^{-2\pi g^2}$
722.	$p^2 e^{-\frac{1}{2}\pi f^2}$	$4\pi\sqrt{2} e^{-2\pi g^2}(4\pi g^2 - 1)$
723.	$p^3 e^{-\frac{1}{2}\pi f^2}$	$-16\pi^2 g\sqrt{2} e^{-2\pi g^2}(4\pi g^2 - 3)$
724.	$p^4 e^{-\frac{1}{2}\pi f^2}$	$16\pi^2\sqrt{2} e^{-2\pi g^2}(16\pi^2 g^4 - 24\pi g^2 + 3)$
725.*	$\frac{1}{p} e^{-\pi\beta f^2}$	$\mathbb{S}_{-1}(g) + \frac{1}{2}\operatorname{erf}(g\sqrt{\pi/\beta}) \mp \frac{1}{2}, \quad 0 < \pm$
726.*	$\frac{1}{p^2} e^{-\pi\beta f^2}$	$\mathbb{S}_{-2}(g) + \frac{1}{2} g \operatorname{erf}(g\sqrt{\pi/\beta}) + \frac{\sqrt{\beta}}{2\pi} e^{-\pi g^2/\beta} \mp \frac{1}{2} g, \quad 0 < \pm$
727.	$\frac{1}{p}\exp(\alpha p^2) - \frac{1}{p}$	$\mp \frac{1}{2}\operatorname{erfc}\frac{\lvert g\rvert}{2\sqrt{\alpha}}, \quad 0 < \pm$
728.	$\frac{1}{p - p_0}\exp[\alpha(p^2 - p_0{}^2)] - \frac{1}{p - p_0}$	$\frac{1}{2}\operatorname{cis}(2\pi f_0 g)\left(\operatorname{erf}\frac{g + 2\alpha p_0}{2\sqrt{\alpha}} \mp 1\right), \quad 0 < \pm$
729.	$\exp(\rho p^2 + \sigma p), \quad 0 < \lvert\rho\rvert$	$\frac{1}{2\sqrt{\pi\rho}}\exp\left[-\frac{(g + \sigma)^2}{4\rho}\right]$
751.	$\sin(ap^2)$	$\frac{1}{2\sqrt{\pi a}}\sin\left(\frac{g^2}{4a} - \frac{\pi}{4}\right)$
752.	$\cos(ap^2)$	$\frac{1}{2\sqrt{\pi a}}\sin\left(\frac{g^2}{4a} + \frac{\pi}{4}\right)$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
753.	$\frac{\sin(ap^2)}{p}$	$\frac{1}{2}\left[S\left(\frac{g}{\sqrt{2\pi a}}\right) - C\left(\frac{g}{\sqrt{2\pi a}}\right)\right]$
754.	$\frac{\cos(ap^2)}{p} - \frac{1}{p}$	$\frac{1}{2}\left[S\left(\frac{g}{\sqrt{2\pi a}}\right) + C\left(\frac{g}{\sqrt{2\pi a}}\right) \mp 1\right]$, $0 < \pm g$
755.	$\frac{\cos(ap^2)}{(p - p_0)\cos(ap_0^2)} - \frac{1}{p - p_0}$	$\frac{\operatorname{cis}(2\pi f_0 g)}{4\cos(ap_0^2)}\left[\exp(iap_0^2)\operatorname{erf}\left(\frac{g}{2\sqrt{ia}} + p_0\sqrt{ia}\right) + \exp(-iap_0^2)\operatorname{erf}\left(\frac{g}{2\sqrt{-ia}} + p_0\sqrt{-ia}\right) \mp 2\cos(ap_0^2)\right]$, $0 < \pm g$
756.	$\frac{\sin(ap^2)}{p^2}$	$\sqrt{\frac{a}{\pi}}\sin\left(\frac{g^2}{4a} + \frac{\pi}{4}\right) + \frac{g}{2}\left[S\left(\frac{g}{\sqrt{2\pi a}}\right) - C\left(\frac{g}{\sqrt{2\pi a}}\right)\right]$
757.	$\frac{\sin(ap^2)}{p^3} - \frac{a}{p}$	$\frac{1}{2}\left[\left(a + \frac{g^2}{2}\right)S\left(\frac{g}{\sqrt{2\pi a}}\right) + \left(a - \frac{g^2}{2}\right)C\left(\frac{g}{\sqrt{2\pi a}}\right) + g\sqrt{\frac{a}{\pi}}\sin\left(\frac{g^2}{4a} + \frac{\pi}{4}\right) \mp a\right]$, $0 < \pm g$
758.	$\sin(ap^2 + \lambda)$	$\frac{1}{2\sqrt{\pi a}}\sin\left(\frac{g^2}{4a} + \lambda - \frac{\pi}{4}\right)$
759.	$\cos(ap^2 + \lambda)$	$\frac{1}{2\sqrt{\pi a}}\sin\left(\frac{g^2}{4a} + \lambda + \frac{\pi}{4}\right)$
760.	$\operatorname{cis}[\pm\pi(f^2 - \frac{1}{8})]$	$\operatorname{cis}[\mp\pi(g^2 - \frac{1}{8})]$
761.	$\cos[\pi(f^2 - \frac{1}{8})]$	$\cos[\pi(g^2 - \frac{1}{8})]$
762.	$\sin[\pi(f^2 - \frac{1}{8})]$	$-\sin[\pi(g^2 - \frac{1}{8})]$

TABLE I (Continued)

Part 8. Other Elementary Transcendental Functions of f.

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
801.	$\exp(-\alpha\sqrt{p})$	$\frac{\alpha}{2g\sqrt{\pi g}}\exp\left(-\frac{\alpha^2}{4g}\right)$, $0 < g$
802.	$p\exp(-\alpha\sqrt{p})$	$\left(\frac{\alpha^2}{2g}-3\right)\frac{\alpha}{4g^2\sqrt{\pi g}}\exp\left(-\frac{\alpha^2}{4g}\right)$, $0 < g$
803.	$\frac{1}{p}[1-\exp(-\alpha\sqrt{p})]$	$\operatorname{erf}\frac{\alpha}{2\sqrt{g}}$, $0 < g$
804.*	$\frac{1}{p^2}[1-\exp(-\alpha\sqrt{p})]+\frac{\alpha^2}{2p}$	$\left(g+\frac{\alpha^2}{2}\right)\operatorname{erf}\frac{\alpha}{2\sqrt{g}}+\alpha\sqrt{\frac{g}{\pi}}\exp\left(-\frac{\alpha^2}{4g}\right)$, $0 < g$
805.	$\frac{1-\exp(-\alpha\sqrt{p})}{p(p+\gamma)}$	$\frac{e^{-\gamma g}}{2\gamma}\left[\exp(-\alpha\sqrt{-\gamma})\operatorname{erfc}\left(\frac{\alpha}{2\sqrt{g}}-\sqrt{-\gamma g}\right)+\exp(\alpha\sqrt{-\gamma})\operatorname{erfc}\left(\frac{\alpha}{2\sqrt{g}}+\sqrt{-\gamma g}\right)\right]+\frac{1}{\gamma}\operatorname{erf}\frac{\alpha}{2\sqrt{g}}-\frac{1}{\gamma}e^{-\gamma g}$, $0 < g$
806.	$\sqrt{p}\exp(-\alpha\sqrt{p})$	$\left(\frac{\alpha^2}{2g}-1\right)\frac{1}{2g\sqrt{\pi g}}\exp\left(-\frac{\alpha^2}{4g}\right)$, $0 < g$
807.	$\frac{1}{\sqrt{p}}\exp(-\alpha\sqrt{p})$	$\frac{1}{\sqrt{\pi g}}\exp\left(-\frac{\alpha^2}{4g}\right)$, $0 < g$
808.*	$\frac{1}{\alpha p\sqrt{p}}\exp(-\alpha\sqrt{p})+\frac{1}{p}$	$\frac{2}{\alpha}\sqrt{\frac{g}{\pi}}\exp\left(-\frac{\alpha^2}{4g}\right)+\operatorname{erf}\frac{\alpha}{2\sqrt{g}}$, $0 < g$
809.	$\frac{\exp(-\alpha\sqrt{p})}{1+\sqrt{\beta p}}$	$\frac{1}{\sqrt{\pi\beta g}}\exp\left(-\frac{\alpha^2}{4g}\right)-\frac{1}{\beta}\exp\frac{\alpha\sqrt{\beta}+g}{\beta}\times\operatorname{erfc}\frac{\alpha\sqrt{\beta}+2g}{2\sqrt{\beta g}}$, $0 < g$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
10.	$\dfrac{p \exp(-\alpha\sqrt{p})}{1+\sqrt{\beta p}}$	$\dfrac{\alpha^2\beta - 2(\beta+\alpha\sqrt{\beta})g + 4g^2}{4\beta g^2\sqrt{\pi\beta g}} \exp\left(-\dfrac{\alpha^2}{4g}\right) - \dfrac{1}{\beta^2} \exp\dfrac{\alpha\sqrt{\beta}+g}{\beta} \operatorname{erfc}\dfrac{\alpha\sqrt{\beta}+2g}{2\sqrt{\beta g}}, \quad 0 < g$
11.	$\dfrac{\exp(-\alpha\sqrt{p})}{p(1+\sqrt{\beta p})} - \dfrac{1}{p}$	$-\operatorname{erf}\dfrac{\alpha}{2\sqrt{g}} - \exp\dfrac{\alpha\sqrt{\beta}+g}{\beta} \times \operatorname{erfc}\dfrac{\alpha\sqrt{\beta}+2g}{2\sqrt{\beta g}}, \quad 0 < g$
12.	$\dfrac{\exp(-\alpha\sqrt{p})}{(p+\gamma)(1+\sqrt{\beta p})}$	$\dfrac{\exp(-\alpha\sqrt{-\gamma}-\gamma g)}{2(1+\sqrt{-\beta\gamma})} \operatorname{erfc}\left(\dfrac{\alpha}{2\sqrt{g}} - \sqrt{-\gamma g}\right) + \dfrac{\exp(\alpha\sqrt{-\gamma}-\gamma g)}{2(1-\sqrt{-\beta\gamma})} \times \operatorname{erfc}\left(\dfrac{\alpha}{2\sqrt{g}} + \sqrt{-\gamma g}\right) - \dfrac{1}{1+\beta\gamma} \exp\dfrac{\alpha\sqrt{\beta}+g}{\beta} \times \operatorname{erfc}\dfrac{\alpha\sqrt{\beta}+2g}{2\sqrt{\beta g}}, \quad 0 < g$
13.	$\dfrac{p \exp(-\alpha\sqrt{p})}{(p+\gamma)(1+\sqrt{\beta p})}$	$\dfrac{-\gamma \exp(-\alpha\sqrt{-\gamma}-\gamma g)}{2(1+\sqrt{-\beta\gamma})} \times \operatorname{erfc}\left(\dfrac{\alpha}{2\sqrt{g}} - \sqrt{-\gamma g}\right) + \dfrac{-\gamma \exp(\alpha\sqrt{-\gamma}-\gamma g)}{2(1-\sqrt{-\beta\gamma})} \times \operatorname{erfc}\left(\dfrac{\alpha}{2\sqrt{g}} + \sqrt{-\gamma g}\right) - \dfrac{1}{\beta(1+\beta\gamma)} \exp\dfrac{\alpha\sqrt{\beta}+g}{\beta} \times \operatorname{erfc}\dfrac{\alpha\sqrt{\beta}+2g}{2\sqrt{\beta g}} + \dfrac{1}{\sqrt{\pi\beta g}} \exp\left(-\dfrac{\alpha^2}{4g}\right), \quad 0 < g$

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
814.	$\frac{\sqrt{p}\exp(-\alpha\sqrt{p})}{1+\sqrt{\beta p}}$	$\frac{\alpha\sqrt{\beta}-2g}{2\beta g\sqrt{\pi g}}\exp\left(-\frac{\alpha^2}{4g}\right)$ $+\frac{1}{\beta\sqrt{\beta}}\exp\frac{\alpha\sqrt{\beta}+g}{\beta}\operatorname{erfc}\frac{\alpha\sqrt{\beta}+2g}{2\sqrt{\beta g}}$, $0<g$
815.	$\frac{\exp(-\alpha\sqrt{p})}{\sqrt{p}(1+\sqrt{\beta p})}$	$\frac{1}{\sqrt{\beta}}\exp\frac{\alpha\sqrt{\beta}+g}{\beta}\operatorname{erfc}\frac{\alpha\sqrt{\beta}+2g}{2\sqrt{\beta g}}$, $0<g$
816.	$\frac{\sqrt{p}\exp(-\alpha\sqrt{p})}{(p+\gamma)(1+\sqrt{\beta p})}$	$\frac{\sqrt{-\gamma}\exp(-\alpha\sqrt{-\gamma}-\gamma g)}{2(1+\sqrt{-\beta\gamma})}$ $\times\operatorname{erfc}\left(\frac{\alpha}{2\sqrt{g}}-\sqrt{-\gamma g}\right)$ $-\frac{\sqrt{-\gamma}\exp(\alpha\sqrt{-\gamma}-\gamma g)}{2(1-\sqrt{-\beta\gamma})}$ $\times\operatorname{erfc}\left(\frac{\alpha}{2\sqrt{g}}+\sqrt{-\gamma g}\right)$ $+\frac{1}{\sqrt{\beta}(1+\beta\gamma)}\exp\frac{\alpha\sqrt{\beta}+g}{\beta}$ $\times\operatorname{erfc}\frac{\alpha\sqrt{\beta}+2g}{2\sqrt{\beta g}}$, $0<g$
817.	$\exp(-\alpha\sqrt{p+\beta})$	$\frac{\alpha}{2g\sqrt{\pi g}}\exp\left(-\beta g-\frac{\alpha^2}{4g}\right)$, $0<g$
818.	$\frac{1}{p}\exp(-\alpha\sqrt{p+\beta}+\alpha\sqrt{\beta})-\frac{1}{p}$	$-\frac{1}{2}\left[\operatorname{erfc}\left(\sqrt{\beta g}-\frac{\alpha}{2\sqrt{g}}\right)\right.$ $\left.-\exp(2\alpha\sqrt{\beta})\operatorname{erfc}\left(\sqrt{\beta g}+\frac{\alpha}{2\sqrt{g}}\right)\right]$, $0<g$

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
319.	$\dfrac{\exp(-\alpha\sqrt{p+\beta})}{p+\gamma}$	$\frac{1}{2}\Big[\exp(-\alpha\sqrt{\beta-\gamma}-\gamma g)\operatorname{erfc}\Big(\dfrac{\alpha}{2\sqrt{g}}-\sqrt{(\beta-\gamma)g}\Big)+\exp(\alpha\sqrt{\beta-\gamma}-\gamma g)\times\operatorname{erfc}\Big(\dfrac{\alpha}{2\sqrt{g}}+\sqrt{(\beta-\gamma)g}\Big)\Big],\quad 0<g$
320.	$\sqrt{p+\beta}\exp(-\alpha\sqrt{p+\beta})$	$\Big(\dfrac{\alpha^2}{2g}-1\Big)\dfrac{1}{2g\sqrt{\pi g}}\exp\Big(-\dfrac{\alpha^2}{4g}-\beta g\Big),\quad 0<g$
321.	$\dfrac{1}{p}\sqrt{\dfrac{p}{\beta}+1}\exp(-\alpha\sqrt{p+\beta}+\alpha\sqrt{\beta})-\dfrac{1}{p}$	$\dfrac{1}{\sqrt{\pi\beta g}}\exp\Big(-\dfrac{\alpha^2}{4g}-\beta g+\alpha\sqrt{\beta}\Big)-\frac{1}{2}\Big[\operatorname{erfc}\Big(\sqrt{\beta g}-\dfrac{\alpha}{2\sqrt{g}}\Big)+\exp(2\alpha\sqrt{\beta})\operatorname{erfc}\Big(\sqrt{\beta g}+\dfrac{\alpha}{2\sqrt{g}}\Big)\Big],\quad 0<g$
322.	$\dfrac{\sqrt{p+\beta}\exp(-\alpha\sqrt{p+\beta})}{p+\gamma}$	$\dfrac{\sqrt{\beta-\gamma}}{2}\Big[\exp(-\alpha\sqrt{\beta-\gamma}-\gamma g)\times\operatorname{erfc}\Big(\dfrac{\alpha}{2\sqrt{g}}-\sqrt{(\beta-\gamma)g}\Big)-\exp(\alpha\sqrt{\beta-\gamma}-\gamma g)\times\operatorname{erfc}\Big(\dfrac{\alpha}{2\sqrt{g}}+\sqrt{(\beta-\gamma)g}\Big)\Big]+\dfrac{1}{\sqrt{\pi g}}\exp\Big(-\dfrac{\alpha^2}{4g}-\beta g\Big),\quad 0<g$
323.	$\dfrac{\exp(-\alpha\sqrt{p+\beta})}{\sqrt{p+\beta}}$	$\dfrac{1}{\sqrt{\pi g}}\exp\Big(-\beta g-\dfrac{\alpha^2}{4g}\Big),\quad 0<g$

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
824.	$\dfrac{\exp\left(-\alpha\sqrt{p+\beta}+\alpha\sqrt{\beta}\right)}{p\sqrt{\dfrac{p}{\beta}+1}}-\dfrac{1}{p}$	$-\dfrac{1}{2}\left[\operatorname{erfc}\left(\sqrt{\beta g}-\dfrac{\alpha}{2\sqrt{g}}\right)+\exp\left(2\alpha\sqrt{\beta}\right)\operatorname{erfc}\left(\sqrt{\beta g}+\dfrac{\alpha}{2\sqrt{g}}\right)\right]$ $0<$
825.	$\dfrac{\exp\left(-\alpha\sqrt{p+\beta}\right)}{(p+\gamma)\sqrt{p+\beta}}$	$\dfrac{1}{2\sqrt{\beta-\gamma}}\left[\exp\left(-\alpha\sqrt{\beta-\gamma}-\gamma g\right)\times\operatorname{erfc}\left(\dfrac{\alpha}{2\sqrt{g}}-\sqrt{(\beta-\gamma)g}\right)-\exp\left(\alpha\sqrt{\beta-\gamma}-\gamma g\right)\times\operatorname{erfc}\left(\dfrac{\alpha}{2\sqrt{g}}+\sqrt{(\beta-\gamma)g}\right)\right],\quad 0<$
841.	$\dfrac{1}{\sqrt{\lvert p\rvert}}\exp\left(-\rho\sqrt{\lvert p\rvert}\right)$	$\sqrt{\dfrac{2}{\pi\lvert g\rvert}}\left[\sin\dfrac{\rho^2}{4\lvert g\rvert}\,C\left(\dfrac{\rho}{\sqrt{2\pi\lvert g\rvert}}\right)-\cos\dfrac{\rho^2}{4\lvert g\rvert}\,S\left(\dfrac{\rho}{\sqrt{2\pi\lvert g\rvert}}\right)\right]+\dfrac{1}{\sqrt{\pi\lvert g\rvert}}\cos\left(\dfrac{\rho^2}{4\lvert g\rvert}+\dfrac{\pi}{4}\right)$
842.	$\dfrac{1}{\sqrt{\lvert p\rvert}}\exp\left(-\rho\sqrt{\lvert p\rvert}-\sigma\lvert p\rvert\right)$	$\dfrac{1}{2\sqrt{\pi(\sigma+ig)}}\exp\dfrac{\rho^2}{4(\sigma+ig)}\operatorname{erfc}\dfrac{\rho}{2\sqrt{\sigma+i}}+\dfrac{1}{2\sqrt{\pi(\sigma-ig)}}\exp\dfrac{\rho^2}{4(\sigma-ig)}\times\operatorname{erfc}\dfrac{\rho}{2\sqrt{\sigma-i}}$
843.*	$\sqrt{\lvert p\rvert}\sin\left(a\sqrt{\lvert p\rvert}\right)$	$-\dfrac{a}{2\pi g^2}-\dfrac{1}{\lvert g\rvert\sqrt{2\pi\lvert g\rvert}}\left[\left(\cos\dfrac{a^2}{4\lvert g\rvert}-\dfrac{a^2}{2\lvert g\rvert}\sin\dfrac{a^2}{4\lvert g\rvert}\right)C\left(\dfrac{a}{\sqrt{2\pi\lvert g\rvert}}\right)+\left(\sin\dfrac{a^2}{4\lvert g\rvert}+\dfrac{a^2}{2\lvert g\rvert}\cos\dfrac{a^2}{4\lvert g\rvert}\right)\times S\left(\dfrac{a}{\sqrt{2\pi\lvert g\rvert}}\right)\right.$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
44.	$\dfrac{\sqrt{\lvert p\rvert}\,\sin(a\sqrt{\lvert p\rvert})}{p}$	$\pm\sqrt{\dfrac{2}{\pi\lvert g\rvert}}\left[\sin\dfrac{a^2}{4\lvert g\rvert}\,S\left(\dfrac{a}{\sqrt{2\pi\lvert g\rvert}}\right)+\cos\dfrac{a^2}{4\lvert g\rvert}\,C\left(\dfrac{a}{\sqrt{2\pi\lvert g\rvert}}\right)\right],\quad 0<\pm g$
45.	$\cos(\alpha\sqrt{\lvert p\rvert})\,e^{-\beta\lvert p\rvert}$	$\dfrac{\beta}{\pi(\beta^2+g^2)}+\dfrac{i\alpha}{4(\beta+ig)\sqrt{\pi(\beta+ig)}}\times\exp\left[-\dfrac{\alpha^2}{4(\beta+ig)}\right]\operatorname{erf}\dfrac{i\alpha}{2\sqrt{\beta+ig}}+\dfrac{i\alpha}{4(\beta-ig)\sqrt{\pi(\beta-ig)}}\times\exp\left[-\dfrac{\alpha^2}{4(\beta-ig)}\right]\operatorname{erf}\dfrac{i\alpha}{2\sqrt{\beta-ig}}$
46.	$\dfrac{\sin(\alpha\sqrt{\lvert p\rvert})\,e^{-\beta\lvert p\rvert}}{\sqrt{\lvert p\rvert}}$	$\dfrac{1}{2i\sqrt{\pi(\beta+ig)}}\exp\left[-\dfrac{\alpha^2}{4(\beta+ig)}\right]\times\operatorname{erf}\dfrac{i\alpha}{2\sqrt{\beta+ig}}+\dfrac{1}{2i\sqrt{\pi(\beta-ig)}}\times\exp\left[-\dfrac{\alpha^2}{4(\beta-ig)}\right]\operatorname{erf}\dfrac{i\alpha}{2\sqrt{\beta-ig}}$
61.	$\dfrac{\exp[-c\sqrt{p(p+\alpha)}]}{\sqrt{p(p+\alpha)}}$	$e^{-\frac{1}{2}\alpha g}I_0\left(\dfrac{\alpha}{2}\sqrt{g^2-c^2}\right),\qquad c<g$
62.*	$\sqrt{\dfrac{p}{p+\alpha}}\exp[-c\sqrt{p(p+\alpha)}]$	$e^{-\frac{1}{2}\alpha c}\mathbb{S}_0(g-c)+\dfrac{\alpha}{2}e^{-\frac{1}{2}\alpha g}\left[\dfrac{g}{\sqrt{g^2-c^2}}I_1\left(\dfrac{\alpha}{2}\sqrt{g^2-c^2}\right)-I_0\left(\dfrac{\alpha}{2}\sqrt{g^2-c^2}\right)\right],\quad c<g$
63.*	$\exp[-c\sqrt{(p+\alpha)(p+\beta)}]$	$e^{-\frac{1}{2}(\alpha+\beta)c}\mathbb{S}_0(g-c)+\dfrac{c(\alpha-\beta)}{2\sqrt{g^2-c^2}}e^{-\frac{1}{2}(\alpha+\beta)g}I_1\left(\dfrac{\alpha-\beta}{2}\times\sqrt{g^2-c^2}\right),\quad c<g$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
864.*	$\sqrt{\frac{p+\alpha}{p+\beta}}\exp\left[-c\sqrt{(p+\alpha)(p+\beta)}\right]$	$e^{-\frac{1}{2}(\alpha+\beta)c}\mathbf{S}_0(g-c) + \frac{\alpha-\beta}{2}e^{-\frac{1}{2}(\alpha+\beta)g}\left[\frac{g}{\sqrt{g^2-c^2}}I_1\left(\frac{\alpha-\beta}{2}\times\sqrt{g^2-c^2}\right)+I_0\left(\frac{\alpha-\beta}{2}\sqrt{g^2-c^2}\right)\right]$ $c <$
865.*	$\exp(-c\sqrt{p^2+a^2})$	$\mathbf{S}_0(g-c) - \frac{ac}{\sqrt{g^2-c^2}}J_1(a\sqrt{g^2-c^2})$, $c <$
866.	$\frac{\exp(-c\sqrt{p^2+a^2})}{\sqrt{p^2+a^2}}$	$J_0(a\sqrt{g^2-c^2})$, $c <$
867.	$\exp(-\alpha\sqrt{\beta^2-p^2})$	$\frac{\alpha\beta K_1(\beta\sqrt{\alpha^2+g^2})}{\pi\sqrt{\alpha^2+g^2}}$
868.	$\frac{\exp(-\alpha\sqrt{\beta^2-p^2})}{\sqrt{\beta^2-p^2}}$	$\frac{1}{\pi}K_0(\beta\sqrt{\alpha^2+g^2})$
869.	$\frac{(\sqrt{p^2+a^2}+p)^{-\nu}}{\sqrt{p^2+a^2}}\exp(-r\sqrt{p^2+a^2})$	$a^{-\nu}\left(\frac{g-r}{g+r}\right)^{\frac{1}{2}\nu}J_\nu(a\sqrt{g^2-r^2})$, $r <$
871.*	$\cos(a\sqrt{\beta^2-p^2})$	$\frac{1}{2}\mathbf{S}_0(g+a)+\frac{1}{2}\mathbf{S}_0(g-a) - \frac{a\beta J_1(\beta\sqrt{a^2-g^2})}{2\sqrt{a^2-g^2}}$, $-a < g <$
872.	$\frac{\sin(a\sqrt{\beta^2-p^2})}{\sqrt{\beta^2-p^2}}$	$\frac{1}{2}J_0(\beta\sqrt{a^2-g^2})$, $-a < g <$
881.	$\tan^{-1}\frac{r}{p+\rho}$	$\frac{1}{g}e^{-\rho g}\sin rg$, $0 <$
891.	$(p+\beta)^{-\alpha}\log(p+\beta)$	$\frac{1}{\Gamma(\alpha)}g^{\alpha-1}e^{-\beta g}[\psi(\alpha)-\log g]$, $0 <$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
92.	$p^{-\alpha}\log p$, $\quad 0 < R(\alpha) < 1$	$\dfrac{\psi(\alpha) - \log g}{\Gamma(\alpha)} g^{\alpha-1}$, $\quad 0 < g$
93.*	$\dfrac{\log p}{p} = \lim\limits_{\beta\to 0} \dfrac{\log (p+\beta)}{p+\beta}$	$\psi(1) - \log g$, $\quad 0 < g$
94.	$\log \dfrac{p+\gamma}{p+\beta}$	$\dfrac{1}{g}(e^{-\beta g} - e^{-\gamma g})$, $\quad 0 < g$
	Part 9. Other Transcendental Functions of f.	
01.	$\exp p^2 \operatorname{erfc} p$	$\dfrac{1}{\sqrt{\pi}} \exp(-\frac{1}{4}g^2)$, $\quad 0 < g$
02.	$\dfrac{1}{\sqrt{p}} \exp\left(\dfrac{1}{p}\right) \operatorname{erfc}\left(\dfrac{1}{\sqrt{p}}\right)$	$\dfrac{1}{\sqrt{\pi g}} \exp(-2\sqrt{g})$, $\quad 0 < g$
11.	$p^{\frac{1}{2}-\alpha} K_{\frac{1}{2}-\alpha}(ap)$, $\quad 0 < R(\alpha) < 1$	$\dfrac{\sqrt{\pi}}{\Gamma(\alpha)} (2a)^{\frac{1}{2}-\alpha} (g^2 - a^2)^{\alpha-1}$, $\quad a < g$
12.	$K_0(ap)$	$\dfrac{1}{\sqrt{g^2 - a^2}}$, $\quad a < g$
13.*	$\dfrac{1}{p} K_0(ap) = \lim\limits_{\beta\to 0} \dfrac{K_0[a(p+\beta)]}{p+\beta}$	$\cosh^{-1}\dfrac{g}{a}$, $\quad a < g$
14.	$p^{\frac{1}{2}-\alpha} I_{\alpha-\frac{1}{2}}(ap)$	$\dfrac{1}{\sqrt{\pi}\Gamma(\alpha)} (2a)^{\frac{1}{2}-\alpha} (a^2 - g^2)^{\alpha-1}$, $\quad -a < g < a$
15.	$(\beta^2 - p^2)^{\frac{1}{2}\nu} K_\nu(\sqrt{\beta^2 - p^2})$	$\dfrac{1}{\sqrt{2\pi}} \left(\dfrac{\beta^2}{g^2+1}\right)^{\frac{1}{2}\nu+\frac{1}{4}} K_{\nu+\frac{1}{2}}(\beta\sqrt{g^2+1})$
16.	$(\rho^2 + f^2)^{-\frac{1}{4}} K_{\frac{1}{4}}(2\pi\rho\sqrt{\rho^2 + f^2})$	$(\rho^2 + g^2)^{-\frac{1}{4}} K_{\frac{1}{4}}(2\pi\rho\sqrt{\rho^2 + g^2})$
17.	$K_0(\beta\sqrt{\rho^2 - p^2})$	$\dfrac{\exp(-\rho\sqrt{g^2+\beta^2})}{2(g^2+\beta^2)^{\frac{1}{2}}}$

* A star marks a pair as being the limit approached by regular pairs.

TABLE I (Continued)

Pair No.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
918.	$K_0(\alpha\lvert p\rvert)$	$\frac{1}{2\sqrt{\alpha^2+g^2}}$
919.	$\lvert p\rvert K_1(\alpha\lvert p\rvert)$	$\frac{\alpha}{2\sqrt{(\alpha^2+g^2)^3}}$
920.	$K_0(\alpha\sqrt{p})$	$\frac{1}{2g}\exp\left(-\frac{\alpha^2}{4g}\right),$ $0 <$
921.	$\sqrt{p}\,K_1(\alpha\sqrt{p})$	$\frac{\alpha}{4g^2}\exp\left(-\frac{\alpha^2}{4g}\right),$ $0 <$
922.	$\frac{\alpha}{\sqrt{p}}K_1(\alpha\sqrt{p})-\frac{1}{p}$	$\exp\left(-\frac{\alpha^2}{4g}\right)-1,$ $0 <$
923.	$\sqrt{p}\exp(p^2)K_{\frac{1}{4}}(p^2)$	$\frac{1}{\sqrt{2g}}\exp(-\frac{1}{8}g^2),$ $0 <$
924.	$\frac{1}{\sqrt{p}}\exp\left(\frac{1}{p}\right)K_{\frac{1}{4}}\left(\frac{1}{p}\right)$	$(2g)^{-\frac{3}{4}}\exp(-2\sqrt{2g}),$ $0 <$
925.	$\frac{1}{\sqrt{p}}\exp\left(-\frac{1}{p}\right)I_{\alpha-\frac{1}{2}}\left(\frac{1}{p}\right)$	$\frac{1}{\sqrt{\pi g}}J_{2\alpha-1}(2\sqrt{2g}),$ $0 <$
931.	$\sqrt{p}K_{\nu+\frac{1}{4}}(p)K_{\nu-\frac{1}{4}}(p),\quad -\frac{3}{4} < R(\nu) < \frac{3}{4}$	$\frac{\sqrt{\pi}(y^{2\nu}+y^{-2\nu})}{\sqrt{2g(g^2-4)}},\ y=\frac{1}{2}(g+\sqrt{g^2-4}),$ $2 <$
932.	$\sqrt{p}\left[I_{\nu-x}(p)I_{-\nu-x}(p)\right]_{x=-\frac{1}{4}}^{x=\frac{1}{4}}$	$\frac{\sqrt{2}(y^{2\nu}+y^{-2\nu})}{\pi\sqrt{\pi g(4-g^2)}},\ y=\frac{1}{2}(g+\sqrt{g^2-4}),$ $0 < g <$
933.	$\sqrt{p}I_v(p)K_{v+\frac{1}{2}}(p)$	$\frac{(-1)^v(x^{2v+\frac{1}{2}}+x^{-2v-\frac{1}{2}})}{\sqrt{2\pi g(4-g^2)}},$ $x=\frac{1}{2}(g+\sqrt{g^2-4}),$ $0 < g <$

TABLE I (Continued)

air o.	Coefficient $F(f)$ for the Cisoidal Oscillation	Coefficient $G(g)$ for the Unit Impulse
4.	$\sqrt{p}\left[J_{-\frac{3}{4}+x+\alpha}(p)\, Y_{-\frac{3}{4}-x+\alpha}(p) \right]_{x=-\frac{1}{4}}^{x=\frac{1}{4}}$	$\dfrac{2^{2\alpha}(g+\sqrt{g^2+4})^{\frac{3}{2}-2\alpha}}{\pi\sqrt{\pi g(g^2+4)}}$, $\quad 0 < g$
5.	$J_{\alpha-\frac{1}{2}}(\sqrt{\delta^2-p^2}+ip)J_{\alpha-\frac{1}{2}}(\sqrt{\delta^2-p^2}-ip)$	$\dfrac{1}{\pi}(4-g^2)^{-\frac{1}{2}}J_{2\alpha-1}(\delta\sqrt{4-g^2})$, $\quad -2 < g < 2$
6.	$I_{\alpha-\frac{1}{2}}(\sqrt{p^2+b^2}-p)K_{\alpha-\frac{1}{2}}(\sqrt{p^2+b^2}+p)$	$(g^2-4)^{-\frac{1}{2}}J_{2\alpha-1}(b\sqrt{g^2-4})$, $\quad 2 < g$
1.*	$\mathbf{S}_0(f) = \dfrac{1}{\pi}\lim_{\beta\to 0}\dfrac{\beta}{\beta^2+f^2}$	1
2.*	$\mathbf{S}_0(f-f_0)$, $\quad$ lim by 981*	$\operatorname{cis}(2\pi f_0 g)$
3.*	$\mathbf{S}_0(f-f_0)+\mathbf{S}_0(f+f_0)$, $\quad$ lim by 981*	$2\cos(2\pi f_0 g)$
4.*	$\mathbf{S}_0(f-f_0)-\mathbf{S}_0(f+f_0)$, $\quad$ lim by 981*	$i2\sin(2\pi f_0 g)$
5.*	$\mathbf{S}_1(f) = \lim_{\beta\to 0}\left(-\dfrac{2\pi f}{\beta\sqrt{\beta}}e^{-\pi f^2/\beta}\right)$	$-i2\pi g$
6.*	$\mathbf{S}_{-1}(f) = \lim_{\beta\to 0}\left(\lambda \pm \dfrac{1}{2}\right)e^{-\beta\lvert f\rvert}$, $\quad 0 < \pm f$	$-\dfrac{1}{i2\pi g}+\lambda\mathbf{S}_0(g)$
7.*	$\mathbf{S}_{-2}(f) = \lim_{\beta\to 0}\left(\dfrac{1}{2}\lvert f\rvert+\lambda f+\mu\right)e^{-\beta\lvert f\rvert}$	$-\dfrac{1}{4\pi^2 g^2}+\dfrac{\lambda}{i2\pi}\mathbf{S}_1(g)+\mu\mathbf{S}_0(g)$

* A star marks a pair as being the limit approached by regular pairs.

TABLE II—ADMITTANCES OF

No.	Admittance $Y(p)$ Illustrative System Cause and Effect	Cause: Unit Impulse $= \mathbb{S}_0(t)$ Effect: $\mathfrak{L}Y(p)$ 403*
1	$Y(p) = \dfrac{Cp + G}{LC(p - p_1)(p - p_2)}$ Inductance L and resistance R in series with parallel combination of capacity C and conductance G. Cause: Voltage across terminals. Effect: Current through network. $\left.\begin{matrix} p_1 \\ p_2 \end{matrix}\right\} = \dfrac{-(RC + LG) \pm \Delta}{2LC}$, $\Delta = \sqrt{(RC - LG)^2 - 4LC}$.	$\dfrac{1}{\Delta}[(Cp_1 + G)e^{p_1 t} - (Cp_2 + G)e^{p_2 t}]$, $0 < t$ 448, 449
2	$Y(p) = Cp + G$ Same as 1, except $R = 0$, $L = 0$.	$C\mathbb{S}_1(t) + G\mathbb{S}_0(t)$ 403*, 404*
3	$Y(p) = \exp\left[-\dfrac{x}{v}\sqrt{(p + \rho)^2 - \sigma^2}\right]$ Semi-infinite smooth line (resistance R, inductance L, conductance G, and capacity C per unit length). Cause: Initial voltage. Effect: Voltage at distance x from end. $v = (LC)^{-\frac{1}{2}}$, $k = (L/C)^{\frac{1}{2}}$, $\alpha = R/(2L)$, $\beta = G/(2C)$, $\rho = \alpha + \beta$, $\sigma = \alpha - \beta$, $z = \sqrt{t^2 - (x/v)^2}$.	$\exp\left(-\dfrac{\rho x}{v}\right)\mathbb{S}_0\left(t - \dfrac{x}{v}\right)$ $+ \dfrac{\sigma x}{vz} e^{-\rho t} I_1(\sigma z)$, $\dfrac{x}{v} < t$ 863*
4	$Y(p) = \dfrac{1}{k}\sqrt{\dfrac{p + \rho - \sigma}{p + \rho + \sigma}} \times \exp\left[-\dfrac{x}{v}\sqrt{(p + \rho)^2 - \sigma^2}\right]$ Same as 3, except Effect: Current at distance x from end.	$\dfrac{1}{k}\exp\left(-\dfrac{\rho x}{v}\right)\mathbb{S}_0\left(t - \dfrac{x}{v}\right)$ $+ \dfrac{1}{k} e^{-\rho t}\left[\dfrac{\sigma t}{z} I_1(\sigma z) - \sigma I_0(\sigma z)\right]$, $\dfrac{x}{v} < t$ 864*

AND TRANSIENTS IN PHYSICAL SYSTEMS

Cause: Unit Step (0, 1) $= \mathbf{S}_{-1}(t)$, $\lambda = \frac{1}{2}$ Effect: $\mathcal{L}[Y(p)/p]$ 415*	Cause: Unit Cisoid $\times$ Unit Step (0, 1) $= e^{p_0 t}\mathbf{S}_{-1}(t)$, $\lambda = \frac{1}{2}$ Effect: $\mathcal{L}[Y(p)/(p - p_0)]$ 440*
$\dfrac{1}{R + G^{-1}} + \dfrac{Cp_1 + G}{\Delta p_1} e^{p_1 t} - \dfrac{Cp_2 + G}{\Delta p_2} e^{p_2 t}, \quad 0 < t$ 448, 454, 415*	$\dfrac{(Cp_0 + G)e^{p_0 t}}{LC(p_0 - p_1)(p_0 - p_2)} + \dfrac{(Cp_1 + G)e^{p_1 t}}{\Delta(p_1 - p_0)} - \dfrac{(Cp_2 + G)e^{p_2 t}}{\Delta(p_2 - p_0)}, \quad 0 < t$ 452, 453
$C\mathbf{S}_0(t) + G, \quad 0 < t$ 403*, 415*	$C\mathbf{S}_0(t) + (Cp_0 + G)e^{p_0 t}, \quad 0 < t$ 438, 441*

TABLE II

No.	Admittance $Y(p)$ Illustrative System Cause and Effect	Cause: Unit Impulse Effect: $\mathcal{L}^{-1}Y(p)$
5	$Y(p) = \exp(-y\sqrt{p+2\beta})$ Same as 3, except $L = 0$. $y = x\sqrt{RC}$. 817	$\frac{y}{2t\sqrt{\pi t}} \exp\left(-\frac{y^2}{4t} - 2\beta t\right), \quad 0 < t$
6	$Y(p) = u\sqrt{p+2\beta}\exp(-y\sqrt{p+2\beta})$ Same as 4, except $L = 0$. $y = x\sqrt{RC}, \quad u = \sqrt{C/R}$. 820	$\frac{u(y^2-2t)}{4t^2\sqrt{\pi t}} \exp\left(-\frac{y^2}{4t} - 2\beta t\right), \; 0 < t$
7	$Y(p) = \frac{\exp(-y\sqrt{p+2\beta})}{u\sqrt{p+2\beta}}$ Same as 3, except $L = 0$ and Cause: Initial current. $y = x\sqrt{RC}, \quad u = \sqrt{C/R}$. 823	$\frac{1}{u\sqrt{\pi t}} \exp\left(-\frac{y^2}{4t} - 2\beta t\right), \quad 0 < t$
8	$Y(p) = \frac{1}{k}\sqrt{\frac{p}{p+2\alpha}} \times \exp\left[-\frac{x}{v}\sqrt{p(p+2\alpha)}\right]$ Same as 4, except $G = 0$. 862*	$\frac{1}{k}\exp\left(-\frac{\alpha x}{v}\right)\mathbb{S}_0\left(t-\frac{x}{v}\right) + \frac{1}{k}e^{-\alpha t}\left[\frac{\alpha t}{z}I_1(\alpha z) - \alpha I_0(\alpha z)\right], \; \frac{x}{v} < t$
9	$Y(p) = k\sqrt{\frac{p+2\alpha}{p}}$ Same as 3, except $G = 0$, $x = 0$, and Cause: Initial current. 553*	$k\mathbb{S}_0(t) + k\alpha e^{-\alpha t}[I_1(\alpha t) + I_0(\alpha t)], \quad 0 < t$

Continued

Cause: Unit Step (0, 1) Effect: $\mathfrak{Sl}[Y(p)/p]$	Cause: Unit Cisoid × Unit Step (0, 1) Effect: $\mathfrak{Sl}[Y(p)/(p-p_0)]$
$\frac{1}{2}\left[\exp(-y\sqrt{2\beta})\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}-\sqrt{2\beta t}\right)\right.$ $\left.+\exp(y\sqrt{2\beta})\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}+\sqrt{2\beta t}\right)\right],$ $0<t$	$\frac{1}{2}e^{p_0 t}\left[\exp(-y\sqrt{2\beta+p_0})\right.$ $\times\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}-\sqrt{(2\beta+p_0)t}\right)$ $\left.+\exp(y\sqrt{2\beta+p_0})\right.$ $\left.\times\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}+\sqrt{(2\beta+p_0)t}\right)\right],$ $0<t$
818, 415*	819
$\frac{u}{\sqrt{\pi t}}\exp\left(-\frac{y^2}{4t}-2\beta t\right)+\frac{u\sqrt{2\beta}}{2}$ $\times\left[\exp(-y\sqrt{2\beta})\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}-\sqrt{2\beta t}\right)\right.$ $\left.-\exp(y\sqrt{2\beta})\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}+\sqrt{2\beta t}\right)\right],$ $0<t$	$\frac{u}{\sqrt{\pi t}}\exp\left(-\frac{y^2}{4t}-2\beta t\right)$ $+\frac{u\sqrt{2\beta+p_0}}{2}e^{p_0 t}\left[\exp(-y\sqrt{2\beta+p_0})\right.$ $\times\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}-\sqrt{(2\beta+p_0)t}\right)$ $\left.-\exp(y\sqrt{2\beta+p_0})\right.$ $\left.\times\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}+\sqrt{(2\beta+p_0)t}\right)\right],$ $0<t$
821, 415*	822
$\frac{1}{2u\sqrt{2\beta}}\left[\exp(-y\sqrt{2\beta})\right.$ $\times\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}-\sqrt{2\beta t}\right)-\exp(y\sqrt{2\beta})$ $\left.\times\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}+\sqrt{2\beta t}\right)\right],$ $0<t$	$\frac{e^{p_0 t}}{2u\sqrt{2\beta+p_0}}\left[\exp(-y\sqrt{2\beta+p_0})\right.$ $\times\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}-\sqrt{(2\beta+p_0)t}\right)$ $-\exp(y\sqrt{2\beta+p_0})$ $\left.\times\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}+\sqrt{(2\beta+p_0)t}\right)\right],$ $0<t$
824, 415*	825
$\frac{1}{k}e^{-\alpha t}I_0(\alpha z),$ $\frac{x}{v}<t$	
861	
$ke^{-\alpha t}[2\alpha t I_1(\alpha t)$ $+(1+2\alpha t)I_0(\alpha t)],$ $0<t$	
554*	

TABLE II

No.	Admittance $Y(p)$ Illustrative System Cause and Effect	Cause: Unit Impulse Effect: $\mathfrak{M}Y(p)$
10	$Y(p) = \exp\left(-\frac{\rho x}{v} - \frac{x}{v}p\right)$ Same as 3, except $R/L = G/C$.	$\exp\left(-\frac{\rho x}{v}\right)\mathbb{S}_0\left(t - \frac{x}{v}\right)$ 601*
11	$Y(p) = \frac{\sqrt{p + 2\alpha}}{\sqrt{p} + \sqrt{p + 2\alpha}}$ Semi-infinite smooth line (resistance R, inductance L, and capacity C per unit length). Cause: Voltage applied through resistance $R_0 = \sqrt{L/C}$. Effect: Voltage at end of line. $\alpha = R/(2L)$.	$\frac{1}{2}\mathbb{S}_0(t) + \frac{1}{2t}e^{-\alpha t}I_1(\alpha t), \qquad 0 < t$ 559*
12	$Y(p) = \frac{\exp(-y\sqrt{p})}{1 + \sqrt{p/\lambda}}$ Semi-infinite smooth line (resistance R and capacity C per unit length). Cause: Voltage applied through resistance R_0. Effect: Voltage at distance x from end. $y = x\sqrt{RC}, \qquad \lambda = R/(CR_0^2)$.	$\sqrt{\frac{\lambda}{\pi t}}\exp\left(-\frac{y^2}{4t}\right) - \lambda\exp(y\sqrt{\lambda} + \lambda t)$ $\times \operatorname{erfc}\left(\frac{y}{2\sqrt{t}} + \sqrt{\lambda t}\right), \qquad 0 < t$ 809
13	$Y(p) = \frac{u\sqrt{p}\exp(-y\sqrt{p})}{1 + \sqrt{p/\lambda}}$ Same as 12, except Effect: Current at distance x from end. $u = \sqrt{C/R}$.	$\frac{u(y - 2t\sqrt{\lambda})}{2t}\sqrt{\frac{\lambda}{\pi t}}\exp\left(-\frac{y^2}{4t}\right)$ $+ u\lambda\sqrt{\lambda}\exp(y\sqrt{\lambda} + \lambda t)$ $\times \operatorname{erfc}\left(\frac{y}{2\sqrt{t}} + \sqrt{\lambda t}\right), \qquad 0 < t$ 814
14	$Y(p) = \frac{u\sqrt{p}}{1 + \sqrt{p/\lambda}}$ Same as 13, except $x = 0$. $u = \sqrt{C/R}$.	$u\sqrt{\lambda}\left[\mathbb{S}_0(t) - \sqrt{\frac{\lambda}{\pi t}} + \lambda e^{\lambda t}\operatorname{erfc}\sqrt{\lambda t}\right], \qquad 0 < t$ 550*

Continued

Cause: Unit Step (0, 1) Effect: $\mathcal{H}[Y(p)/p]$	Cause: Unit Cisoid × Unit Step (0, 1) Effect: $\mathcal{H}[Y(p)/(p-p_0)]$
$\exp\left(-\frac{\rho x}{v}\right)$, $\frac{x}{v} < t$ 602*	$\exp\left[-\frac{x}{v}(\rho + p_0) + p_0 t\right]$, $\frac{x}{v} < t$ 604
$1 - \frac{1}{2}e^{-\alpha t}[I_0(\alpha t) + I_1(\alpha t)]$, $0 < t$ 560, 415*	
$\operatorname{erfc}\frac{y}{2\sqrt{t}} - \exp(y\sqrt{\lambda} + \lambda t) \times \operatorname{erfc}\left(\frac{y}{2\sqrt{t}} + \sqrt{\lambda t}\right)$, $0 < t$ 811, 415*	$\frac{1}{2}e^{p_0 t}\left[\frac{\exp(-y\sqrt{p_0})}{1+\sqrt{p_0/\lambda}}\operatorname{erfc}\left(\frac{y}{2\sqrt{t}} - \sqrt{p_0 t}\right) + \frac{\exp(y\sqrt{p_0})}{1-\sqrt{p_0/\lambda}}\operatorname{erfc}\left(\frac{y}{2\sqrt{t}} + \sqrt{p_0 t}\right)\right] - \frac{1}{1-p_0/\lambda}\exp(y\sqrt{\lambda} + \lambda t) \times \operatorname{erfc}\left(\frac{y}{2\sqrt{t}} + \sqrt{\lambda t}\right)$, $0 < t$ 812
$u\sqrt{\lambda}\exp(y\sqrt{\lambda} + \lambda t)\operatorname{erfc}\left(\frac{y}{2\sqrt{t}} + \sqrt{\lambda t}\right)$, $0 < t$ 815	$\frac{u\sqrt{p_0}}{2}e^{p_0 t}\left[\frac{\exp(-y\sqrt{p_0})}{1+\sqrt{p_0/\lambda}} \times \operatorname{erfc}\left(\frac{y}{2\sqrt{t}} - \sqrt{p_0 t}\right) - \frac{\exp(y\sqrt{p_0})}{1-\sqrt{p_0/\lambda}} \times \operatorname{erfc}\left(\frac{y}{2\sqrt{t}} + \sqrt{p_0 t}\right)\right] + \frac{u\sqrt{\lambda}}{1-p_0/\lambda} \times \exp(y\sqrt{\lambda} + \lambda t)\operatorname{erfc}\left(\frac{y}{2\sqrt{t}} + \sqrt{\lambda t}\right)$, $0 < t$ 816
$u\sqrt{\lambda}e^{\lambda t}\operatorname{erfc}\sqrt{\lambda t}$, $0 < t$ 551	$\frac{u\sqrt{\lambda}}{1-p_0/\lambda}\left[\sqrt{\frac{p_0}{\lambda}}e^{p_0 t}\operatorname{erf}\sqrt{p_0 t} - \frac{p_0}{\lambda}e^{p_0 t} + e^{\lambda t}\operatorname{erfc}\sqrt{\lambda t}\right]$, $0 < t$ 552

TABLE II

No.	Admittance $Y(p)$ Illustrative System Cause and Effect	Cause: Unit Impulse Effect: $\mathcal{H}Y(p)$
15	$Y(p) = \dfrac{C_0 p \exp(-y\sqrt{p})}{1 + \sqrt{p/\mu}}$ Semi-infinite smooth line (resistance R and capacity C per unit length). Cause: Voltage applied through capacity C_0. Effect: Current at distance x from end. $y = x\sqrt{RC}, \qquad \mu = C/(RC_0^2)$.	$\dfrac{C_0(y^2 - 2yt\sqrt{\mu} - 2t + 4\mu t^2)}{4t^2}\sqrt{\dfrac{\mu}{\pi t}}$ $\times \exp\left(-\dfrac{y^2}{4t}\right)$ $- C_0\mu^2 \exp(y\sqrt{\mu} + \mu t)$ $\times \operatorname{erfc}\left(\dfrac{y}{2\sqrt{t}} + \sqrt{\mu t}\right), \qquad 0 < t$ 810
16	$Y(p) = \dfrac{C_0 p}{1 + \sqrt{p/\mu}}$ Same as 15, except $x = 0$.	$- C_0\mu\mathfrak{S}_0(t) + \dfrac{C_0(2\mu t - 1)}{2t}\sqrt{\dfrac{\mu}{\pi t}}$ $- C_0\mu^2 e^{\mu t} \operatorname{erfc}\sqrt{\mu t}, \qquad 0 < t$ 544*
17	$Y(p) = \dfrac{w^{2n+1}[\sqrt{(p+\lambda)^2 + w^2} + (p+\lambda)]^{-2n}}{k\sqrt{(p+\lambda)^2 + w^2}}$ Semi-infinite artificial line (series element: resistance R and inductance L; shunt element: conductance G and capacity C; $R/L = G/C$; mid-series termination). Cause: Applied voltage. Effect: Current in nth section. $k = (L/C)^{\frac{1}{2}}$, $\lambda = R/L = G/C$, $w = 2(LC)^{-\frac{1}{2}}$.	$\dfrac{2}{L} e^{-\lambda t} J_{2n}(wt), \qquad 0 < t$ 575
18	$Y(p) = \dfrac{2(2\alpha)^n}{R}\sqrt{\dfrac{p}{p+2\alpha}}$ $\times (\sqrt{p+2\alpha} + \sqrt{p})^{-2n}$ Semi-infinite artificial line (series element: resistance R; shunt element: capacity C; mid-series termination). Cause: Applied voltage. Effect: Current in nth section. $\alpha = 2/(RC)$.	$\dfrac{\alpha}{R} e^{-\alpha t}[I_{n-1}(\alpha t) - 2I_n(\alpha t) + I_{n+1}(\alpha t)],$ $0 < t$ 573

Continued

Cause: Unit Step (0, 1) Effect: $\mathcal{H}[Y(p)/p]$	Cause: Unit Cisoid × Unit Step (0, 1) Effect: $\mathcal{H}[Y(p)/(p-p_0)]$
$C_0\sqrt{\frac{\mu}{\pi t}}\exp\left(-\frac{y^2}{4t}\right)$ $-C_0\mu\exp(y\sqrt{\mu}+\mu t)\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}+\sqrt{\mu t}\right),\quad 0<t$ 809	$\frac{1}{2}C_0p_0\,e^{p_0t}\left[\frac{\exp(-y\sqrt{p_0})}{1+\sqrt{p_0/\mu}}\times\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}-\sqrt{p_0t}\right)\right.$ $\left.+\frac{\exp(y\sqrt{p_0})}{1-\sqrt{p_0/\mu}}\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}+\sqrt{p_0t}\right)\right]$ $+C_0\sqrt{\frac{\mu}{\pi t}}\exp\left(-\frac{y^2}{4t}\right)-\frac{C_0\mu}{1-p_0/\mu}$ $\times\exp(y\sqrt{\mu}+\mu t)\operatorname{erfc}\left(\frac{y}{2\sqrt{t}}+\sqrt{\mu t}\right),\quad 0<t$ 813
$C_0\sqrt{\frac{\mu}{\pi t}}-C_0\mu e^{\mu t}\operatorname{erfc}\sqrt{\mu t},\quad 0<t$ 543	$C_0\sqrt{\frac{\mu}{\pi t}}+\frac{C_0\mu}{1-p_0/\mu}\left[\frac{p_0}{\mu}e^{p_0t}-\frac{p_0}{\mu}\sqrt{\frac{p_0}{\mu}}e^{p_0t}\right.$ $\left.\times\operatorname{erf}\sqrt{p_0t}-e^{\mu t}\operatorname{erfc}\sqrt{\mu t}\right],\quad 0<t$ 545
$\frac{2}{R}e^{-\alpha t}I_n(\alpha t),\quad 0<t$ 574	

TABLE II

No.	Admittance $Y(p)$ Illustrative System Cause and Effect	Cause: Unit Impulse Effect: $\mathfrak{L}^{-1} Y(p)$
19	$Y(p) = \exp(x - \lvert x\rvert \sqrt{p^2+1})$ Vertical atmospheric waves, axis of x vertically upwards, velocity = 1, height of homogeneous atmosphere = $\frac{1}{2}$. Cause: Vertical displacement at $x = 0$. Effect: Vertical displacement at time t of particle whose undisturbed position is x.	$e^x \mathfrak{S}_0(t - \lvert x\rvert) - \frac{\lvert x\rvert e^x}{\sqrt{t^2 - x^2}} \times J_1(\sqrt{t^2 - x^2}), \quad \lvert x\rvert < t$ 865*
20	$Y(p) = \frac{\exp(x - \lvert x\rvert \sqrt{p^2+1})}{2\sqrt{p^2+1}}$ Same as 19, except Cause: Vertical force at $x = 0$.	$\frac{e^x}{2} J_0(\sqrt{t^2 - x^2}), \quad \lvert x\rvert < t$ 866
21	$Y(p) = \frac{\lvert y\rvert}{\pi r} \sqrt{p}\, K_1(r\sqrt{p})$ Flow of heat in infinite plane. Cause: Temperature impulse at origin, temperature maintained zero along x-axis, except at origin. Effect: Temperature of point with coördinates (x, y) at time t. $r = \sqrt{x^2 + y^2}$.	$\frac{\lvert y\rvert}{4\pi t^2} \exp\left(-\frac{r^2}{4t}\right), \quad 0 < t$ 921
22	$Y(p) = \frac{1}{p}\left[1 - \exp\left(-y\sqrt{\frac{p}{\nu}}\right)\right]$ Horizontal oscillations of deep viscous fluid, axis of y vertical, bottom plane $y = 0$, kinematic coefficient of viscosity = ν. Cause: Applied horizontal force. Effect: Displacement of particle at y at time t, y assumed small.	$\operatorname{erf} \frac{y}{2\sqrt{\nu t}}, \quad 0 < t$ 803
23	$Y(p) = \frac{1}{2\pi} K_0\left(\frac{rp}{c}\right)$ Water waves radiating from center in an unlimited sheet of uniform depth h, gravity constant = g. Cause: Pressure at the origin. Effect: Velocity potential at distance r, time t. $c^2 = gh$.	$\frac{1}{2\pi} \cdot \frac{1}{\sqrt{t^2 - \frac{r^2}{c^2}}}, \quad \frac{r}{c} < t$ 912

Continued

Cause: Unit Step (0, 1) Effect: $\mathfrak{Sli}[Y(p)/p]$	Cause: Unit Cisoid × Unit Step (0, 1) Effect: $\mathfrak{Sli}[Y(p)/(p - p_0)]$
$\frac{\lvert y\rvert}{\pi r^2} \exp\left(-\frac{r^2}{4t}\right)$, $0 < t$	
922, 415*	
$y\sqrt{\frac{t}{\nu\pi}} \exp\left(-\frac{y^2}{4\nu t}\right) - \frac{y^2}{2\nu} + \left(\frac{y^2}{2\nu} + t\right) \operatorname{erf} \frac{y}{2\sqrt{\nu t}}$, $0 < t$	$\frac{1}{p_0} e^{p_0 t} - \frac{1}{p_0} \operatorname{erf} \frac{y}{2\sqrt{\nu t}} - \frac{e^{p_0 t}}{2p_0}\left[\exp\left(-y\sqrt{\frac{p_0}{\nu}}\right) \times \operatorname{erfc}\left(\frac{y}{2\sqrt{\nu t}} - \sqrt{p_0 t}\right) + \exp\left(y\sqrt{\frac{p_0}{\nu}}\right) \operatorname{erfc}\left(\frac{y}{2\sqrt{\nu t}} + \sqrt{p_0 t}\right)\right]$,
804*, 415*	805 $0 < t$
$\frac{1}{2\pi} \cosh^{-1} \frac{ct}{r}$, $\frac{r}{c} < t$	
913*	

TABLE II

No.	Admittance $Y(p)$ Illustrative System Cause and Effect	Cause: Unit Impulse Effect: $\mathcal{M}Y(p)$
24	$Y(p) = \dfrac{\sin[(\pi - y)p]}{\sin \pi p}$ Flow of electricity in thin plane infinite strip, axis of x along lower edge of strip, axis of y across, width of strip $= \pi$, upper edge ($y = \pi$) maintained at zero potential. Cause: Potential along x axis. Effect: Potential at point (x, y).	$\dfrac{1}{2\pi} \cdot \dfrac{\sin y}{\cosh x - \cos y}$ 615
25	$Y(p) = \dfrac{\cos[(\pi - y)p]}{\cos \pi p}$ Same as 24, except upper edge ($y = \pi$) is insulated.	$\dfrac{1}{\pi} \cdot \dfrac{\sin \frac{1}{2}y \cosh \frac{1}{2}x}{\cosh x - \cos y}$ 616
26	$Y(p) = \exp(\kappa t p^2)$ Linear flow of heat in infinite solid, diffusivity κ, axis of x in direction of flow. Cause: Initial temperature. Effect: Temperature at time t at point x.	$\dfrac{1}{2\sqrt{\pi \kappa t}} \exp\left(-\dfrac{x^2}{4\kappa t}\right)$ 701
27	$Y(p) = \cos(tp^2)$ Transverse oscillations of infinite elastic plate; x and y axes in the plate, but all points with same y coördinate have same displacement. Cause: Initial displacement. Effect: Displacement perpendicular to plate at time t of point whose coördinate is x.	$\dfrac{1}{2\sqrt{\pi t}} \sin\left(\dfrac{x^2}{4t} + \dfrac{\pi}{4}\right)$ 752

Continued

Cause: Unit Step $(-\frac{1}{2}, +\frac{1}{2})$ Effect: $\mathcal{M}[Y(p)/p]$	Cause: Unit Cisoid × Unit Step $(-\frac{1}{2}, +\frac{1}{2})$ Effect: $\mathcal{M}[Y(p)/(p-p_0)]$
$\frac{1}{\pi}\tan^{-1}\frac{\tanh\frac{1}{2}x}{\tan\frac{1}{2}y}$ 617, 415*	
$\frac{1}{\pi}\tan^{-1}\frac{\sinh\frac{1}{2}x}{\sin\frac{1}{2}y}$ 618, 415*	
$\frac{1}{2}\operatorname{erf}\frac{x}{2\sqrt{\kappa t}}$ 727, 415*	$\frac{1}{2}\exp(\kappa t p_0^2 + p_0 x)\operatorname{erf}\left(\frac{x}{2\sqrt{\kappa t}} + p_0\sqrt{\kappa t}\right)$ 728, 440*
$\frac{1}{2}\left[S\left(\frac{x}{\sqrt{2\pi t}}\right) + C\left(\frac{x}{\sqrt{2\pi t}}\right)\right]$ 754, 415*	$\frac{1}{4}\exp(p_0 x + i t p_0^2)\operatorname{erf}\left(\frac{x}{2\sqrt{it}} + p_0\sqrt{it}\right)$ $+ \frac{1}{4}\exp(p_0 x - i t p_0^2)$ $\times \operatorname{erf}\left(\frac{x}{2\sqrt{-it}} + p_0\sqrt{-it}\right)$ 755, 440*

TABLE II

No.	Admittance $Y(p)$ Illustrative System Cause and Effect	Cause: Unit Impulse Effect: $\mathfrak{I}\bar{l}\bar{l}\, Y(p)$
28	$Y(p) = \dfrac{\sin(tp^2)}{p^2}$ Same as 27, except Cause: Initial velocity.	$\sqrt{\dfrac{t}{\pi}} \sin\left(\dfrac{x^2}{4t} + \dfrac{\pi}{4}\right) + \dfrac{x}{2}\left[S\left(\dfrac{x}{\sqrt{2\pi t}}\right) - C\left(\dfrac{x}{\sqrt{2\pi t}}\right)\right]$ 756
29	$Y(p) = \cos(t\sqrt{1-p^2})$ Same as 19, except Cause: Initial displacement multiplied by e^{-x}. Effect: Vertical displacement multiplied by e^{-x} at time t of particle whose undisturbed position is x.	$\frac{1}{2}[\mathbf{S}_0(x-t) + \mathbf{S}_0(x+t)] - \dfrac{tJ_1(\sqrt{t^2-x^2})}{2\sqrt{t^2-x^2}}, \quad -t < x < t$ 871*
30	$Y(p) = \dfrac{\sin(t\sqrt{1-p^2})}{\sqrt{1-p^2}}$ Same as 29, except Cause: Initial velocity multiplied by e^{-x}.	$\frac{1}{2}J_0(\sqrt{t^2-x^2}), \quad -t < x < t$ 872
31	$Y(p) = e^{-\lvert yp\rvert}$ Flow of electricity in infinite thin plane, x and y axes in the plane. Cause: Potential along x axis. Effect: Potential at point (x, y).	$\dfrac{1}{\pi} \cdot \dfrac{\lvert y\rvert}{x^2+y^2}$ 632
32	$Y(p) = \cosh(atp)$ Transverse motion of infinite stretched elastic string, axis of x along equilibrium position of string, velocity of propagation along string $= a$. Cause: Initial displacement. Effect: Normal displacement of particle at x at time t.	$\frac{1}{2}[\mathbf{S}_0(x-at) + \mathbf{S}_0(x+at)]$ 619*

Continued

Cause: Unit Step $(-\frac{1}{2}, +\frac{1}{2})$ Effect: $\mathfrak{Sli}[Y(p)/p]$	Cause: Unit Cisoid $\times$ Unit Step $(-\frac{1}{2}, +\frac{1}{2})$ Effect: $\mathfrak{Sli}[Y(p)/(p-p_0)]$
$\frac{1}{2}\left[\left(t+\frac{x^2}{2}\right)S\left(\frac{x}{\sqrt{2\pi t}}\right)+\left(t-\frac{x^2}{2}\right)\times C\left(\frac{x}{\sqrt{2\pi t}}\right)+x\sqrt{\frac{t}{\pi}}\sin\left(\frac{x^2}{4t}+\frac{\pi}{4}\right)\right]$ 757, 415*	
$\frac{1}{\pi}\tan^{-1}\frac{x}{\lvert y\rvert}$ 633, 415*	
$\pm\frac{1}{2}$, $\quad at<\pm x$ 620, 415*	$\begin{cases}\pm\frac{1}{2}e^{p_0x}\cosh(atp_0), & at<\pm x\\ \frac{1}{2}e^{p_0x}\sinh(atp_0), & -at<x<at\end{cases}$ 621, 440*

TABLE II

No.	Admittance $Y(p)$ Illustrative System Cause and Effect	Cause: Unit Impulse Effect: $\mathfrak{L}^{-1}Y(p)$
33	$Y(p) = \dfrac{\sinh(atp)}{ap}$ Same as 32, except Cause: Initial velocity.	$\dfrac{1}{2a}, \qquad -at < x < at$ 622
34	$Y(p) = \dfrac{1}{\rho}\cos(\alpha\sqrt{\lvert p\rvert})e^{y\lvert p\rvert}$ Waves on deep water, axis of y vertically upwards, axis of x in the surface, density $= \rho$, gravity constant $= g$, $y \leqq 0$. Cause: Initial surface-impulse along x axis. Effect: Velocity potential at time t at point (x, y). $h = \dfrac{\alpha}{\sqrt{2\pi\lvert x\rvert}}, \qquad \alpha = t\sqrt{g}.$	$\dfrac{-y}{\pi\rho(x^2+y^2)} + \dfrac{i\alpha}{4\rho\sqrt{\pi}(-y+ix)^{3/2}} \times \exp\left[-\dfrac{\alpha^2}{4(-y+ix)}\right] \times \operatorname{erf}\dfrac{i\alpha}{2\sqrt{-y+ix}} + \dfrac{i\alpha}{4\rho\sqrt{\pi}(-y-ix)^{3/2}} \times \exp\left[-\dfrac{\alpha^2}{4(-y-ix)}\right] \times \operatorname{erf}\dfrac{i\alpha}{2\sqrt{-y-ix}}$ 845
35	$Y(p) = -\dfrac{\sqrt{\lvert p\rvert}}{\rho\sqrt{g}}\sin(\alpha\sqrt{\lvert p\rvert})$ Same as 34, except Effect: Surface elevation at time t at point x.	$\dfrac{\alpha}{2\pi\rho x^2\sqrt{g}} + \dfrac{1}{\rho\lvert x\rvert\sqrt{2\pi g\lvert x\rvert}} \times \{[\cos(\tfrac{1}{2}\pi h^2) - \pi h^2\sin(\tfrac{1}{2}\pi h^2)]C(h) + [\sin(\tfrac{1}{2}\pi h^2) + \pi h^2\cos(\tfrac{1}{2}\pi h^2)]S(h)\}$ 843*
36	$Y(p) = \sqrt{g}\dfrac{\sin(\alpha\sqrt{\lvert p\rvert})}{\sqrt{\lvert p\rvert}}e^{y\lvert p\rvert}$ Same as 34, except Cause: Initial surface elevation.	$\dfrac{\sqrt{g}}{2i\sqrt{\pi(-y+ix)}} \times \exp\left[-\dfrac{\alpha^2}{4(-y+ix)}\right] \times \operatorname{erf}\dfrac{i\alpha}{2\sqrt{-y+ix}} + \dfrac{\sqrt{g}}{2i\sqrt{\pi(-y-ix)}} \times \exp\left[-\dfrac{\alpha^2}{4(-y-ix)}\right] \times \operatorname{erf}\dfrac{i\alpha}{2\sqrt{-y-ix}}$ 846

Continued

Cause: Unit Step $(-\frac{1}{2}, +\frac{1}{2})$ Effect: $\mathcal{H}[Y(p)/p]$	Cause: Unit Cisoid × Unit Step $(-\frac{1}{2}, +\frac{1}{2})$ Effect: $\mathcal{H}[Y(p)/(p-p_0)]$
$\begin{cases} \pm\frac{1}{2}t, & at < \pm x \\ \dfrac{x}{2a}, & -at < x < at \end{cases}$ 623, 415*	$\begin{cases} \pm\dfrac{1}{2ap_0}e^{p_0x}\sinh(atp_0), & at < \pm x \\ \dfrac{1}{2ap_0}[e^{p_0x}\cosh(atp_0) - 1], & -at < x < at \end{cases}$ 624, 440*
$\mp\dfrac{1}{\rho}\sqrt{\dfrac{2}{\pi g\lvert x\rvert}}[\sin(\frac{1}{2}\pi h^2)S(h) + \cos(\frac{1}{2}\pi h^2)C(h)], \quad 0 < \pm x$ 844	

TABLE II

No.	Admittance $Y(p_1, p_2)$ Illustrative System Cause and Effect	Cause: Unit Impulse Effect: $\partial/\partial_1 \partial/\partial_2 Y(p_1, p_2)$
37	$Y(p_1, p_2) = \cos [t(p_1^2 + p_2^2)]$ Transverse oscillations of infinite elastic plate, x and y axes in the plate. Cause: Initial displacement. Effect: Displacement perpendicular to plate at time t of point whose coördinates are x and y.	$\frac{1}{4\pi t} \sin \frac{x^2 + y^2}{4t}$ 759, 758
38	$Y(p_1, p_2) = \exp (- z\sqrt{\lvert p_1^2 + p_2^2 \rvert})$ Velocity potential function in semi-infinite incompressible fluid, x and y axes in surface of fluid, z extending down, $z \geqq 0$. Cause: Velocity potential at surface, $z = 0$. Effect: Velocity potential at point (x, y, z).	$\frac{1}{2\pi} \cdot \frac{z}{(x^2 + y^2 + z^2)^{3/2}}$ 867, 919
39	$Y(p_1, p_2) = \frac{\exp (- z\sqrt{\lvert p_1^2 + p_2^2 \rvert})}{\sqrt{\lvert p_1^2 + p_2^2 \rvert}}$ Newtonian potential function in semi-infinite solid, x and y axes in face of solid, z extending into solid, $z \geqq 0$. Cause: Normal potential derivative at surface, $z = 0$. Effect: Potential at point (x, y, z).	$\frac{1}{2\pi} \cdot \frac{1}{\sqrt{x^2 + y^2 + z^2}}$ 868, 918

Continued

Cause: Unit Step $(-\frac{1}{2}, +\frac{1}{2})$ Effect: $\mathfrak{M}_1\mathfrak{M}_2[Y(p_1, p_2)/(p_1p_2)]$	Cause: Unit Cisoid $\times$ Unit Step $(-\frac{1}{2}, +\frac{1}{2})$ Effect: $\mathfrak{M}_1\mathfrak{M}_2 \dfrac{Y(p_1, p_2)}{(p_1-p_0)(p_2-p_0)}$
$\frac{1}{2}\left[S\left(\frac{x}{\sqrt{2\pi t}}\right) C\left(\frac{y}{\sqrt{2\pi t}}\right) + C\left(\frac{x}{\sqrt{2\pi t}}\right) S\left(\frac{y}{\sqrt{2\pi t}}\right)\right]$ 753; 754, 415*	
$\frac{1}{2\pi} \tan^{-1} \frac{xy}{z\sqrt{x^2+y^2+z^2}}$ †	
$\frac{x}{4\pi} \log \frac{\sqrt{x^2+y^2+z^2}+y}{\sqrt{x^2+y^2+z^2}-y} + \frac{y}{4\pi} \log \frac{\sqrt{x^2+y^2+z^2}+x}{\sqrt{x^2+y^2+z^2}-x} - \frac{z}{2\pi} \tan^{-1} \frac{xy}{z\sqrt{x^2+y^2+z^2}}$ †	

† This solution was obtained by double integration of the unit impulse solution, not by the operation indicated at the head of the column. The two pairs required for this operation nave not yet been found in closed form.

Automatic Machine Gaging

By C. W. ROBBINS

NOTE: This paper discusses the advantages to be gained in certain types of large scale production by the substitution of automatic machine gaging for hand testing. For testing carbon protector blocks, a machine has been developed which accepts all blocks in case a certain dimension lies between 0.0024" and 0.0032" and rejects those when the dimension is 0.0023" or less or 0.0033" or more. This machine will effect a saving of $8000 per year over the cost of hand gaging on an output of 4,500,000 blocks. The saving effected by another recently developed machine replacing a manual test is approximately $1200 a year on a production of 2,500,000 pieces, but a far more important consideration than this money saving is the elimination of an operation so monotonous that it was difficult to keep any operator on it for more than a brief period. The author points out that in some instances automatic machine gaging of the entire product will cost less than a sampling inspection in which there must be included in the direct cost of inspection the cost of some additional supervision and control.

THE cost of testing and gaging parts manufactured in large quantities frequently warrants the construction of special machinery for this work which may be more or less automatic in operation. At the Hawthorne Works of the Western Electric Company considerable study has been given to the problem during the past ten or twelve years and several such machines have been developed. The work has recently assumed more important proportions and many important developments have materialized in the last two or three years.

Some of these machines perform a single operation while others perform several operations successively. Some are automatically fed from a hopper; others are fed by an operator, who at the same time performs some visual operation. Usually each type of piece to be gaged forms a distinct problem, and a single paper to be most useful can be suggestive only as to the procedure to be followed and methods that may be used. To this end it seems best to describe with considerable detail some of the machines that are in successful operation.

Single Test Machine, Automatic Feed

A single purpose gaging machine with automatic feed is shown in Figs. 1 and 2. The part to be tested, shown in Fig. 2 (*a*), is used in the construction of switchboard plugs. It consists of a piece of 5/32 in. brass tubing, $1\frac{1}{2}$ in. long, having a sleeve soldered at one end and a plug soldered in the other. The machine applies a 50 pound test to the two soldered joints simultaneously.

Within the hopper *H*, Fig. 1, a shaft having three slotted arms is

revolved slowly through the ratchet R. The slotted arms pick up the parts suspended in the slot by the sleeve and deliver them into the chute A, Figs. 1 and 2. An intermittently revolving turret B, Fig. 2,

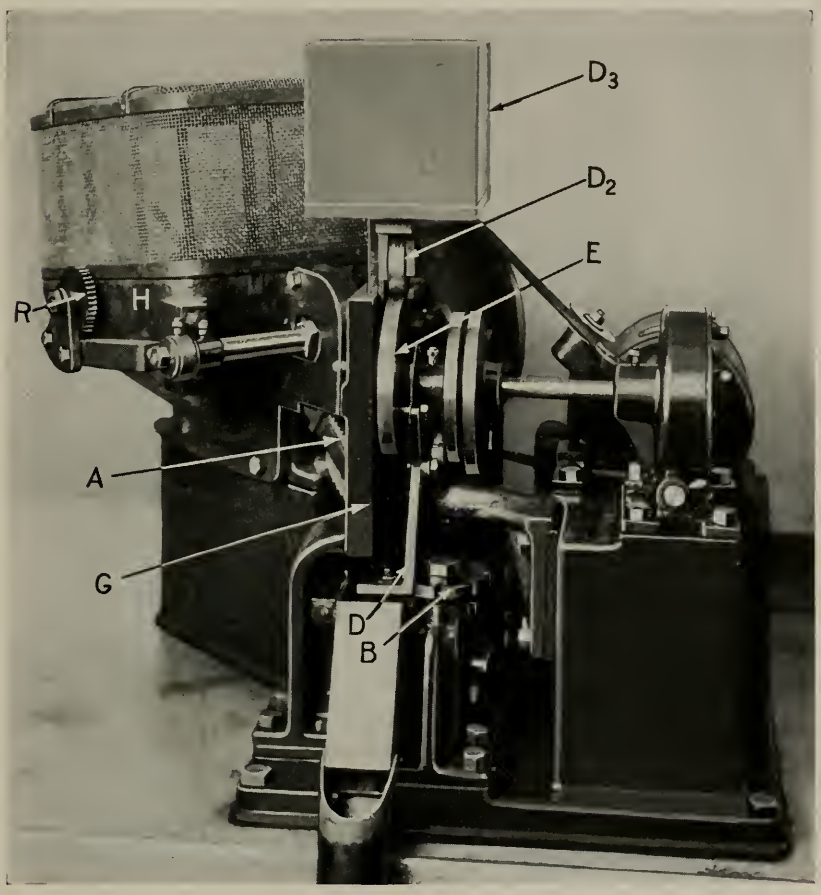


Fig. 1

having four notches or chucks, receives the parts from the chute and carries them under the plunger D, attached to the slide D_1, which carries the 50 pound weight D_3. The slide with the weight is raised and lowered at the proper time by the cam E acting on the roller D_2 attached to the slide.

After the turret has carried the part to the testing position and stopped, the plunger D enters the tube, gradually applying the load (furnished by the weight D_3) to the soldered joints X and Y, Fig. 2.

If either joint is defective, the weight and slide will not be supported but will drop to the limit allowed by the cam at E_1. This permits the wedge-shaped piece D_4, mounted on the slide D_1, to push the far side of the lever F (pivoted to the base at C_3) to the right. The near side of F

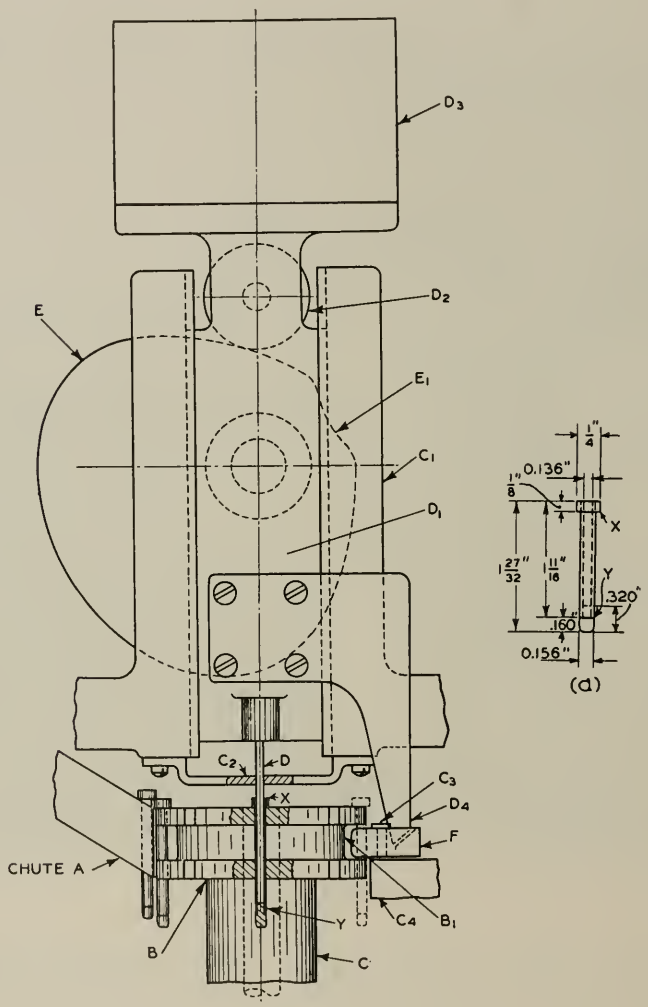


Fig. 2

is pushed into the groove B, and when the turret turns the defective piece is ejected. The lever F is reset by a cam after each operation. If the soldering is good, the load is supported by the piece for a short time while the roller D_2 is passing across the depression E_1, after which

the cam raises the weight, the turret turns and the piece is discharged into the O.K. chute after passing the lever *F*.

The saving effected by the automatic machine replacing the manual test is approximately $1,200 per year on a production of 2,500,000

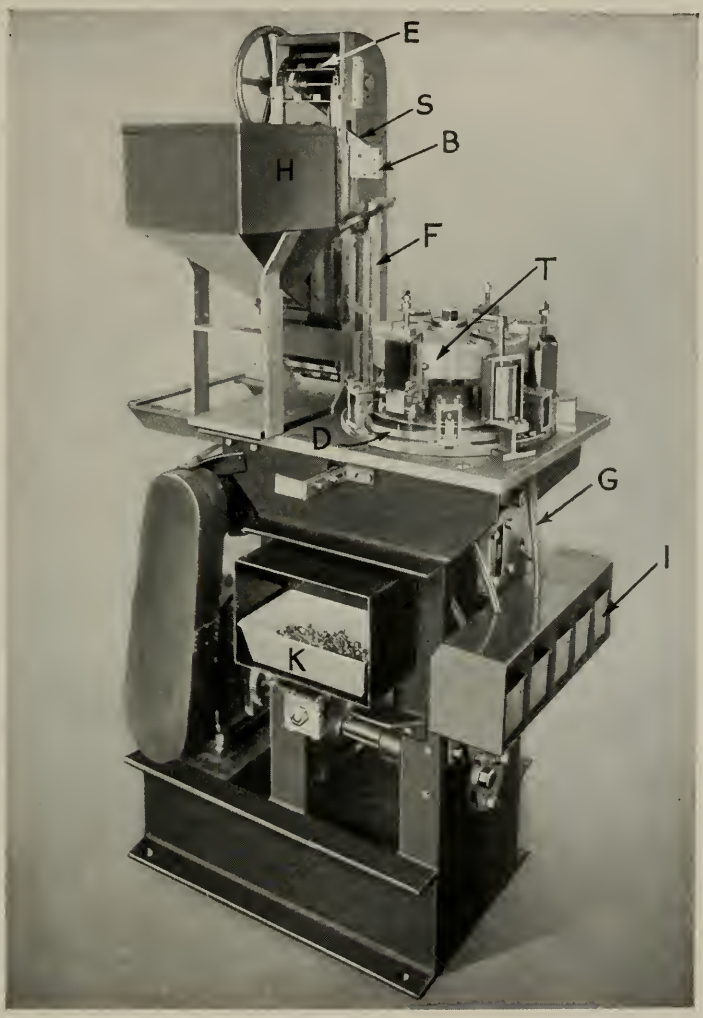


Fig. 3

pieces. However, by far the more important consideration is the elimination of an operation that was so monotonous and tiresome as to make it very difficult to keep any operator on it for more than a brief period.

Multi-Test Machine, Automatic Feed

An automatic gaging machine for applying four tests to a piece of telephone apparatus is shown in Fig. 3.

The part tested, shown in Fig. 4 (*a*), is a heat coil used to protect telephone exchange equipment against excessive electrical currents that may accidentally come in over the line wires. It consists of a tiny coil wound around a copper sleeve, into the end of which sleeve is

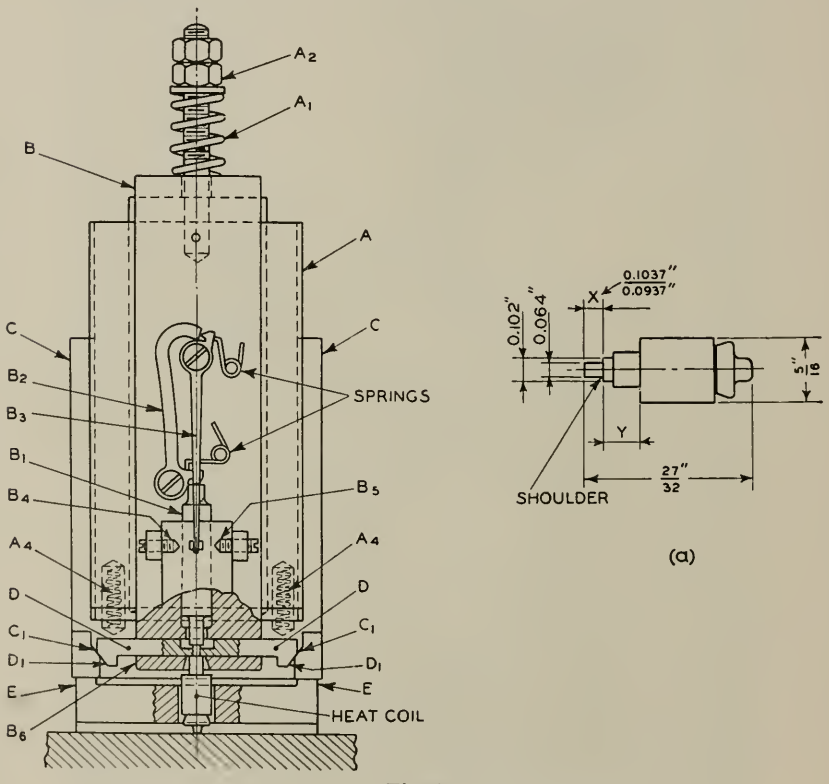


Fig. 4

soldered with low melting point solder a projecting pin. An excessive current through the coil melts the solder, allowing a contact spring to press the pin into the sleeve, which movement of the contact spring opens the circuit.

The machine gages the length of the pin X, the external length of sleeve Y, tests the strength of the soldered joint and measures the electrical resistance of the coil for high and low limits.

Referring to Fig. 3, there is an intermittently rotating disc D fitted

with twelve chucks for holding the parts to be tested and a vertically reciprocating turret head *T* which carries the gages and contact fixtures for making the tests.

The hopper *H*, chain elevator *E* and feed tube *F* are shown in more detail in Fig. 5. The coils are carried up from the hopper by the

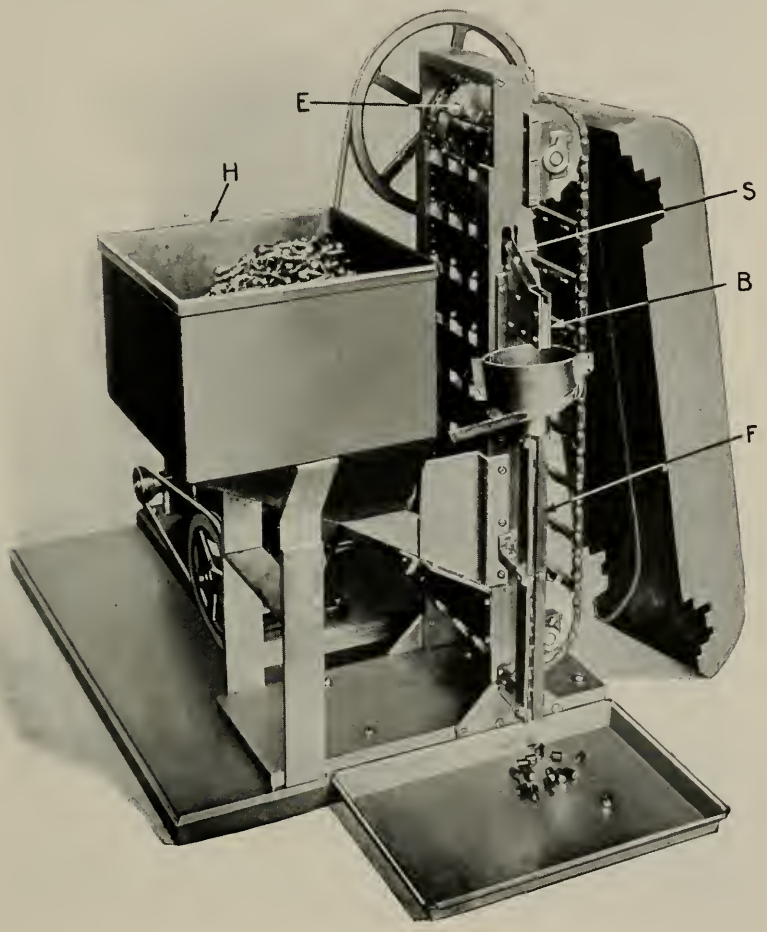


Fig. 5

elevator, two on each cross bar, and drop one after the other into the sloping chute *S*. Since the parts must be right end up for testing, the turning device *B* (shown in detail in Fig. 6) is placed at the end of the chute to turn over those pieces that are not already right end up. From the turning device the parts drop into the vertical feed tube *F*.

The chucks on the intermittently rotating disc take them one at a time from the feed tube and carry them under the gaging heads.

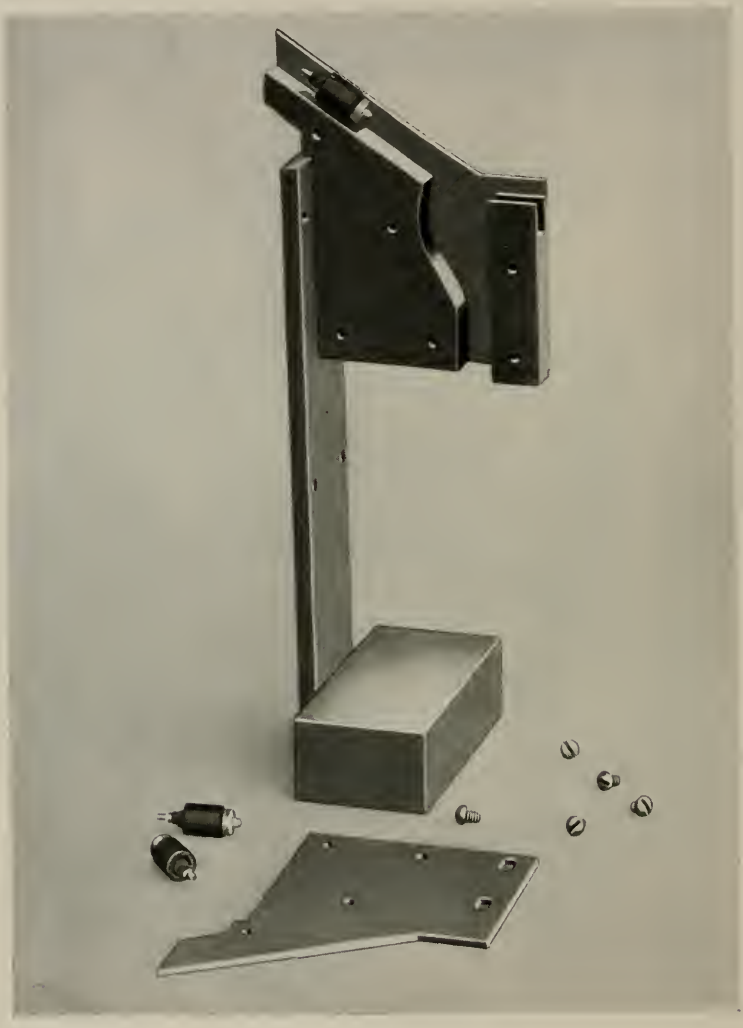


Fig. 6

In the first position the chuck picks a coil from the feed tube. In positions 2, 4, 6, 8 and 10 the coil is tested respectively for right end up, length of pin X, low limit of resistance, high limit of resistance and length of sleeve Y. At positions 3, 5, 7, 9 and 11 are located electromagnets, each controlled by the testing device in the position preceding

it. These are for the purpose of ejecting defective parts so that if a part fails to meet the test at any position an electrical contact is closed which through the electro-magnets sets a trip at the next succeeding position of the chuck, and when the defective part reaches this position it is ejected from the chuck and through one of the tubes G falls into the proper compartment of the container I. At the 12th position the good parts are released from the chuck and fall into the pan K.

A multiple lever gage for this class of work is shown in Fig. 4, which shows the gage for length of pin X and also tests the strength of a

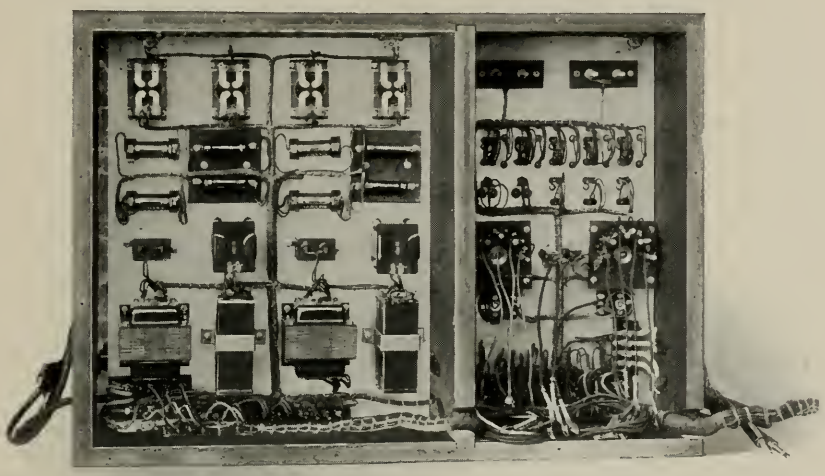


Fig. 7

soldered joint between pin X and sleeve Y. A is the body carrying two sliding members B and C. B carries the gaging mechanism and electrical contacts. B_6 is a preliminary centering guide for the heat coil. Centering slides D, D carried in B are normally held open by springs (not shown). A is normally lifted in C by springs A_4. Slide B is normally held down on A by means of spring A_1.

As the entire gage (A, B, C, D) descends, the heat coil is centered approximately by B_6. Then slide C is restrained by an anvil E, and as A continues downward the slides D, D are closed by the beveled surfaces C_1D_1, thereby centering coil and providing the gaging surface for the shoulder formed by the sleeve. As A continues downward, the gaging surfaces on D, D engage the shoulder and the pressure for operating the gage mechanism is transmitted from A to B through

spring A_1, which limits the testing pressure applied to the soldered joint.

If the length of pin is within limits, the electrical circuit remains open and the coil passes on to the next test. If it is too short, the circuit is closed through lever B_3 coming in contact with B_4, and if too long the contact is made with B_5.

If the soldered joint fails, the effect is the same as a short pin. In either case the tripping magnet operates and at the next position of

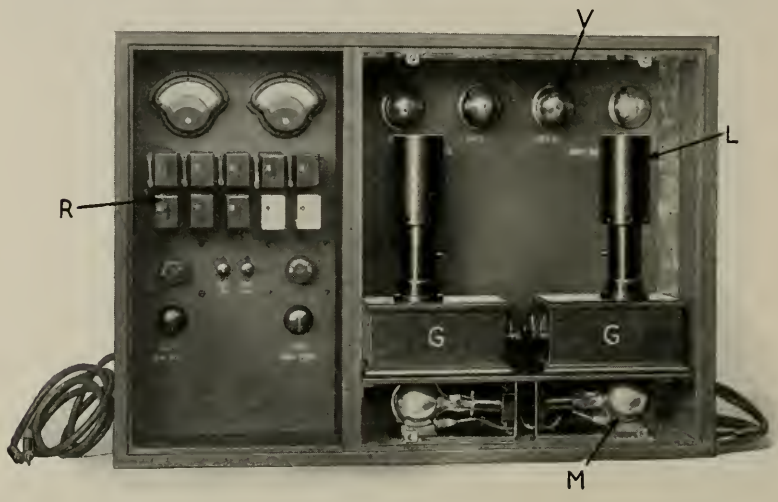


Fig. 8

the disc the defective part is released and drops into the proper container.

For testing the electrical resistance, contact is made with the coil terminals by the gaging machine and the resistance measurement proper is made by means of the apparatus shown in Figs. 7 and 8, which is essentially two Wheatstone resistance bridges, one for checking the resistance of the coil against a low limit and the other against a high limit. The galvanometers G, for indicating the balance of the bridges, each have a small rectangular, delicately pivoted coil which rotates between the pole piece of a strong magnet. The end of a long pointer attached to the coil is broadened out and contains a narrow slot which, in connection with a fixed slot, forms a shutter that passes or intercepts (depending on the position of the coil) a beam of light from a small lamp in the hood L passing to the photo-electric cell M.

The photo-electric cell is connected in the circuit of a vacuum tube amplifier, the tubes of which are shown at V. The position of the small shutter on the galvanometer needle is determined by the relative

value of the resistance of the coil under test to that of standards contained in the bridge. If this resistance is too high in one case or too low in the other, the shutter is closed, preventing the light beam from reaching the photo-electric cell. This in turn through the action

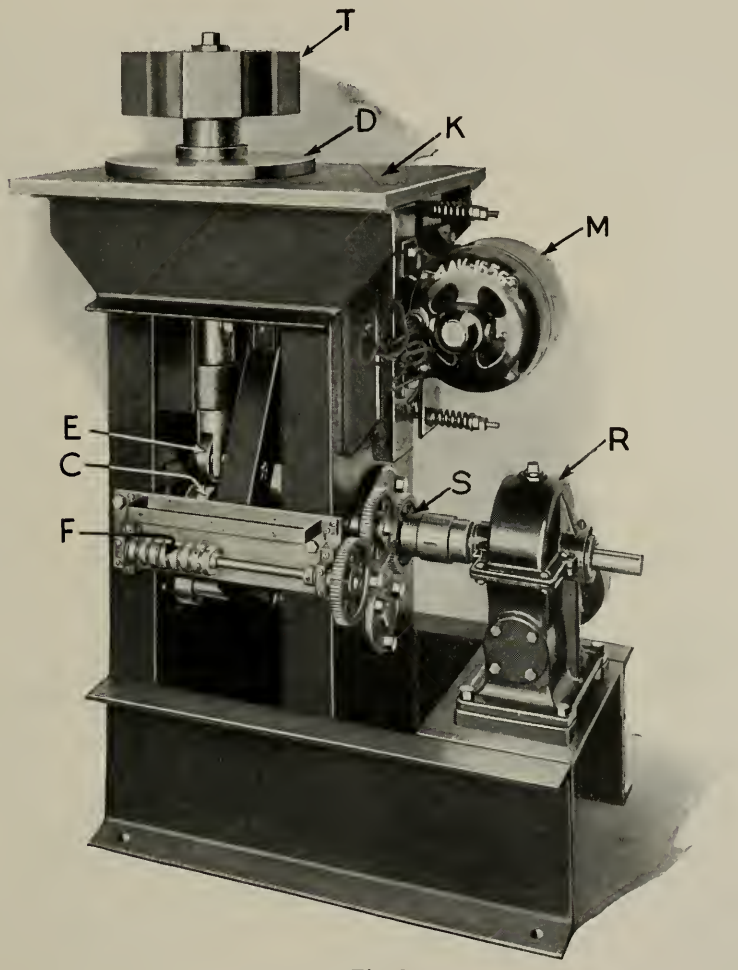


Fig. 9

of the vacuum tube amplifier and relays *R* actuates the trip on the machine which discharges the coil at the proper point.

While designing this machine much attention was given to producing a type that could be adapted readily to the testing of other parts requiring several operations.

Fig. 9 illustrates the fundamental parts of the machine. It is

individually driven by the motor *M* belted to the reducing gear *R* which is attached to the main shaft *S*. The turntable *D* is given an intermittent motion by a Geneva gear and the turret head *T* has a vertical reciprocating motion from the cam *C* on the main shaft through the roller *E*. The cams *F* are used for operating a series of electrical contacts which work in synchronism with the other parts of the machine controlling the sequence of testing operations and the disposition of parts.

The frame is built of welded structural steel. The turntable may be fitted with a variety of chucks or holding fixtures and the turret with various forms of gages or testing apparatus. The cams and gear ratios may be changed to accommodate a wide range of testing requirements. Space is provided at *K* for a hopper or other feeding device. This type of machine is suitable for multiple tests on parts requiring special holding fixtures.

Multi-Unit Testing Machine—Semi-Automatic Feed

An entirely different type of gaging and testing machine is shown in Figs. 10 and 11, which, as illustrated, is equipped for testing porcelain

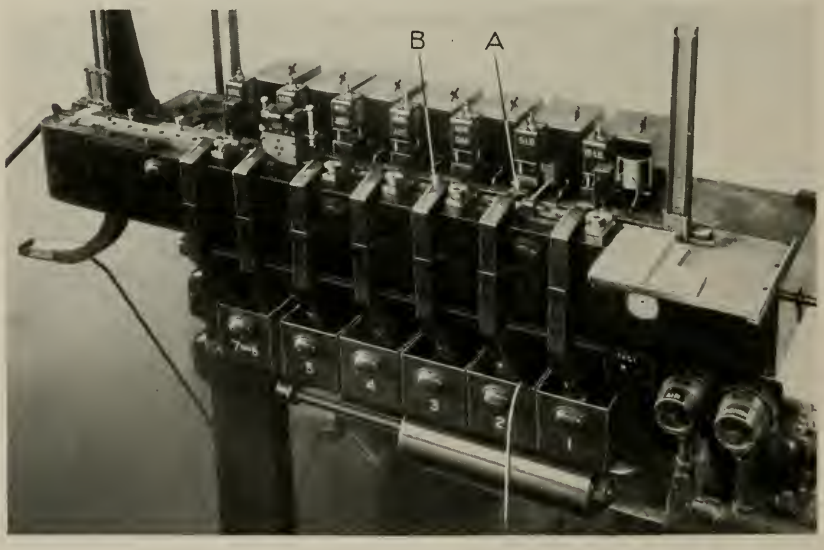


Fig. 10

protector blocks used for protecting telephone apparatus against high voltage electrical currents or static discharges.

These are porcelain blocks $1\frac{1}{4}$ in. x $\frac{3}{8}$ in. x $\frac{9}{32}$ in., having a recessed

surface into the center of which is inserted a small carbon block having a face 0.0370 in. x 0.110 in., cemented in with a low melting point glass. The face of the carbon block is underflush with the rim of the porcelain block, Fig. 12 (*a*).

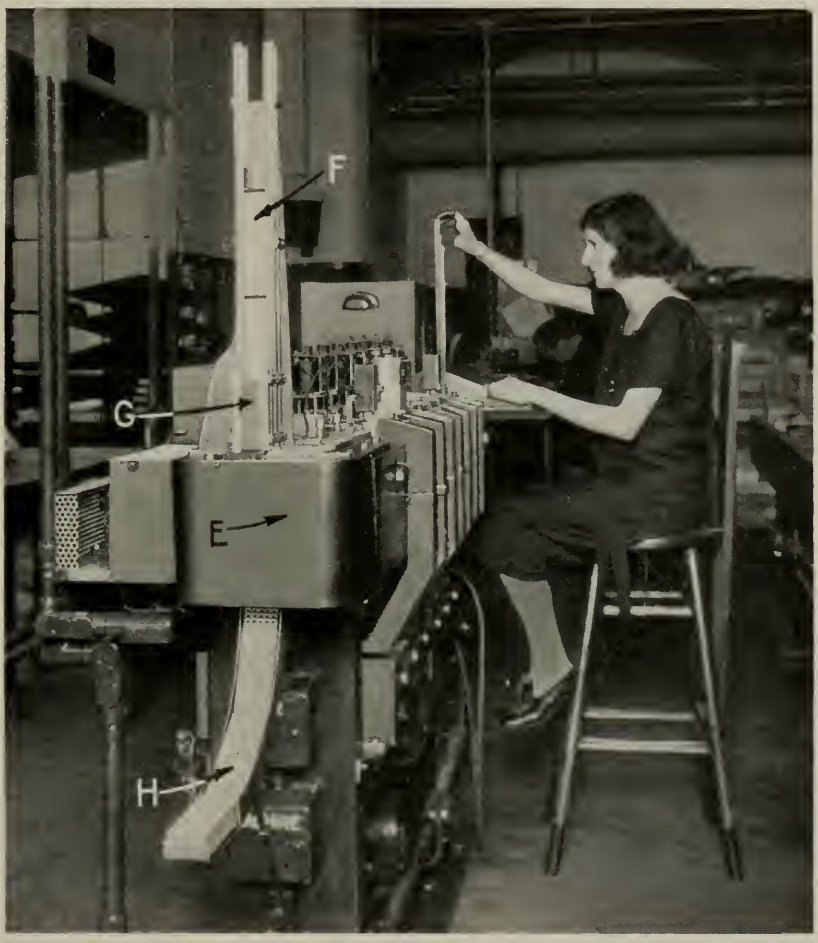
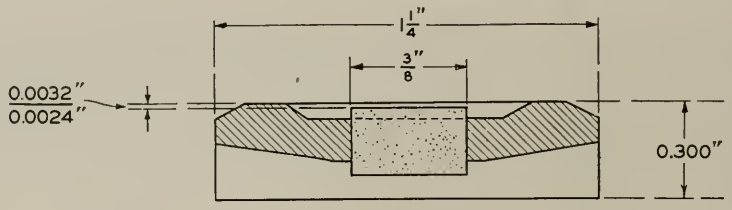


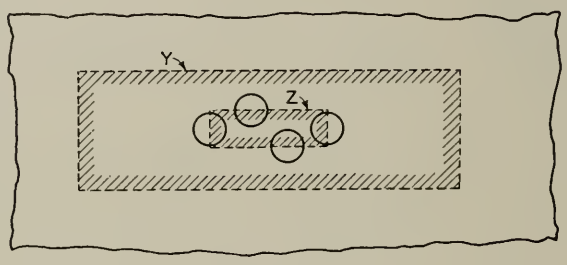
Fig. 11

As shown in Fig. 11, the blocks are being stacked by hand in the top of a vertical chute from which they are automatically fed to the machine at the bottom, but the feeding arrangement shown separately in Fig. 13 is now being added. This consists of a rotating disc having two V-shaped grooves in the surface and above it a stationary plate having a spiral slot. The operator places the blocks rear side up (by

sense of touch) against the front side of the central opening of the stationary plate, three blocks being shown in this position at A. The disc carries them into the spiral slot at B and while they are passing the front opening in the top plate at C they are given a visual inspection.



(a)



(b)

Fig. 12

During the second round they are turned over by the action of the two V grooves in the rotating disc and the radial motion given them by the spiral slot, and the front face is turned up for visual inspection during the second passage across the front opening at D. Any with visible defects are picked off by hand, while the others passing on through the last turn of the spiral at E are fed into the machine (Fig. 10) for the following gaging operations:

1. 15 lb. weight test for defective cementing.
2. 5 lb. weight test to detect misplaced inserts.
3. Gage height of back face of insert—minimum 0.046 in.
4. Gage thickness of block—maximum 0.220 in.
5. Gage thickness of block—minimum 0.205 in.
6. Gage the underflush dimension of insert which must be maximum 0.0032 in., minimum 0.0024 in., for at least half the area of the face of the carbon insert.
7. Gage the underflush dimension for minimum 0.0024 in. over entire face of insert.

The gage for operation No. 6 has four $\frac{1}{8}$ in. plungers (*P*, Fig. 15) arranged to make contact with the insert as shown at (*b*), Fig. 12, and the electrical contacts of the gage controlled by the plungers are connected to a bank of relays so that if any three or all four of the gage points are within the limits the block is passed, but if two or more are outside the limits the block is rejected. This gage, shown in Figs. 14, 15, 16, and 17, is without pivots, the moving parts being controlled

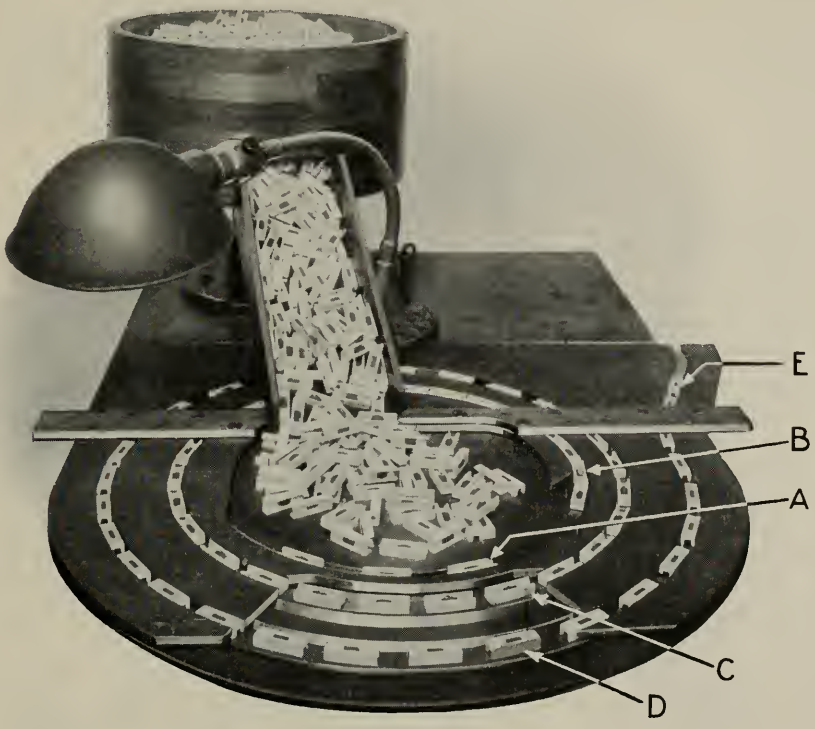


Fig. 13

by thin steel reeds as shown in Fig. 14. Fig. 15 shows a partial, and Fig. 16 a complete, assembly, while Fig. 17 shows the equalizing levers for centering the block in the gage. Using master steel gage blocks, the contact points are adjusted by the screws *A*, Fig. 16, to accept blocks if the underflush dimension is minimum 0.0024 in. and maximum 0.0032 in., and reject blocks when the dimension is 0.0023 in. or less or 0.0033 in. or more.

The single plunger gages used for operations 2, 3, 4, 5 and 7 are also of the reed type and are similar to that shown in Fig. 18. These have a sliding electrical contact instead of a point contact, the pointer *A*

sliding on the surface of the insulated block *B* and making contact with the flush metal insert *C*.

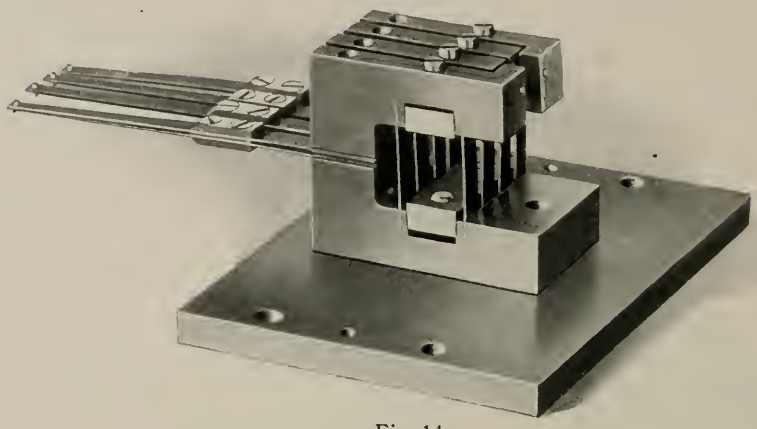


Fig. 14

Considerable experience has been gained in the design and use of the reed type gages and they are proving very satisfactory for a wide variety of uses. They are relatively inexpensive to build, require but

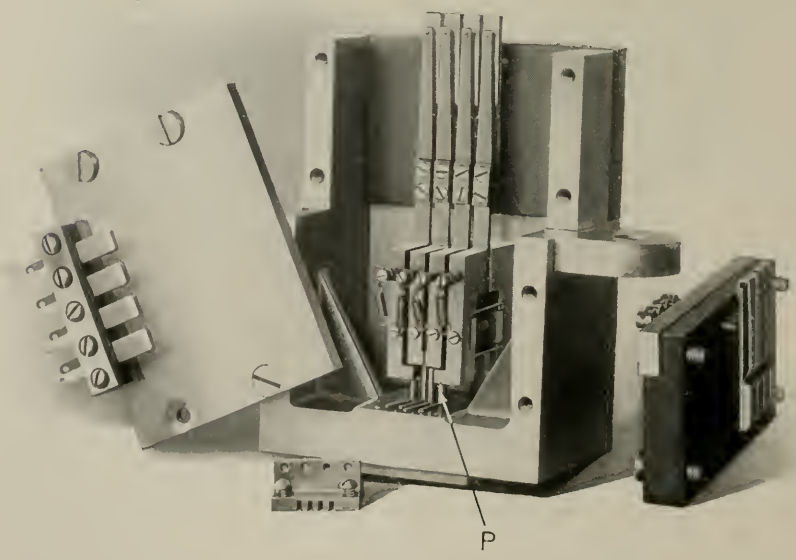


Fig. 15

little maintenance, the pressure on the gaging point may be kept low if desired and they are reliable in action.

Following each gaging operation the blocks pass between air blast tips *A*, Fig. 10, and the square tubes *B* marked 1 to 7 leading to receptacles bearing the same numbers. The electrical contact in each gage, which is closed by a defective block, sets a trip connected to the adjacent air tube so that, as the block passes, the air cock is opened for an instant and the defective block is blown out of the test line into the tube, through which it falls into the proper container. A small air compressor is installed as part of the machine.

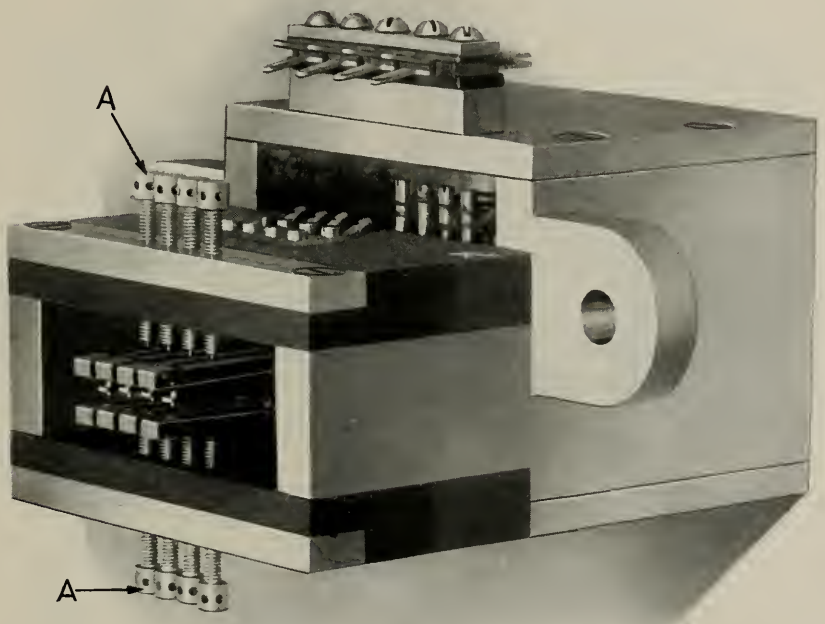


Fig. 16

The O.K. blocks pass along to the automatic packing attachment *E* (shown in more detail in Fig. 11), which places 100 of them in a box in layers of five each with cardboard separators between layers. The empty boxes are shown in the magazine *F*, the separators at *G*, and the filled boxes emerging at *H*.

The feed table shown in Fig. 13 will be placed at the same end of the machine as the packing attachment and the blocks will be carried

from the feed table to the far end of the machine by a conveyor. By this arrangement one operator can feed the machine, make the visual inspection and remove the finished packages. This machine will effect a saving of $8,000 per year over the cost of hand gaging methods

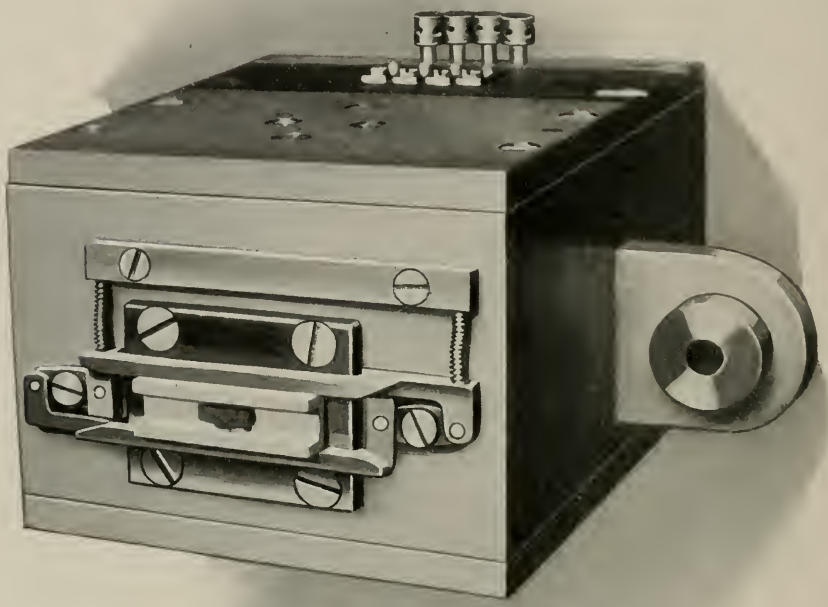


Fig. 17

on an output of 4,500,000 blocks. A similar machine is being built for another size of blocks.

Each gage with its associated equipment is an independent unit as shown in Fig. 19. The gages are located at *G*, either above or below the working surface. Relays and other electrical apparatus at *E*. The solenoid for opening the air cock is shown at *S* and the air blast tip at *T*. The connection for the electrical supply for each unit is made with the cord and plug *P*. The cams *C* and contact springs *D* operate in synchronism with the other parts of the machine and control details of the gaging operation and the air blast.

The main shaft of the machine can be seen at *A*, which drives the disc *B* through worm gears. When the units are attached the pin *K* engages a slot on a disc attached to the rear of the unit shaft *I*. Attached to the front end of the same shaft are two eccentrics *H* (shown in Fig. 20) which operate the feeding device located just back of the blocks shown in the illustration, Figs. 10 and 11.

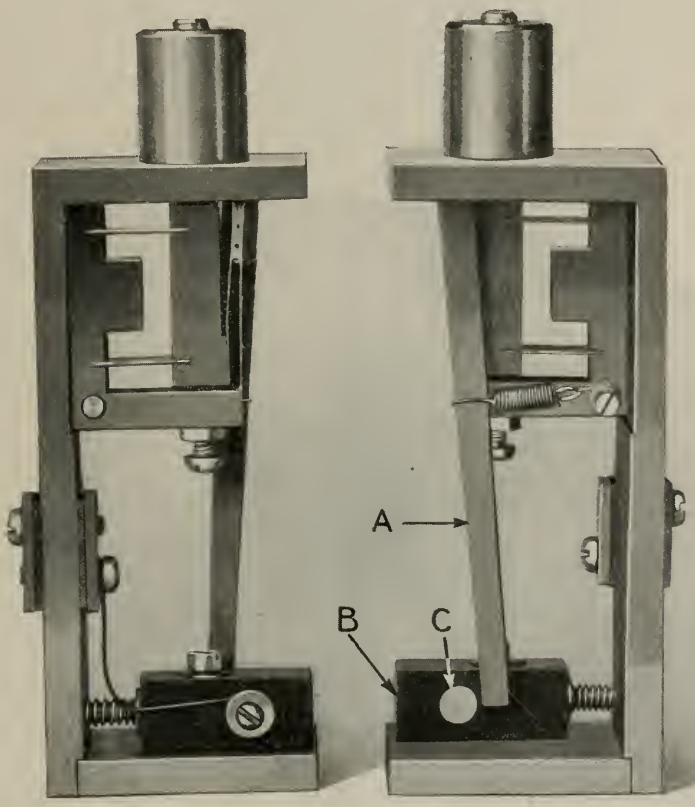


Fig. 18

The feeding mechanism (Fig. 20) is of the finger bar type consisting of two reciprocating bars *A* and *B*, both having the same travel.

Fig. 20 indicates the relative position of the driving parts. Finger bar *A* is operated by a bell crank gear segment and eccentric properly timed to step the protector blocks to the gaging position immediately prior to the gaging operation.

Feed fingers *C* are so located in this finger bar that the protector blocks are centrally located under the gage when the finger bar comes to rest at its forward position.

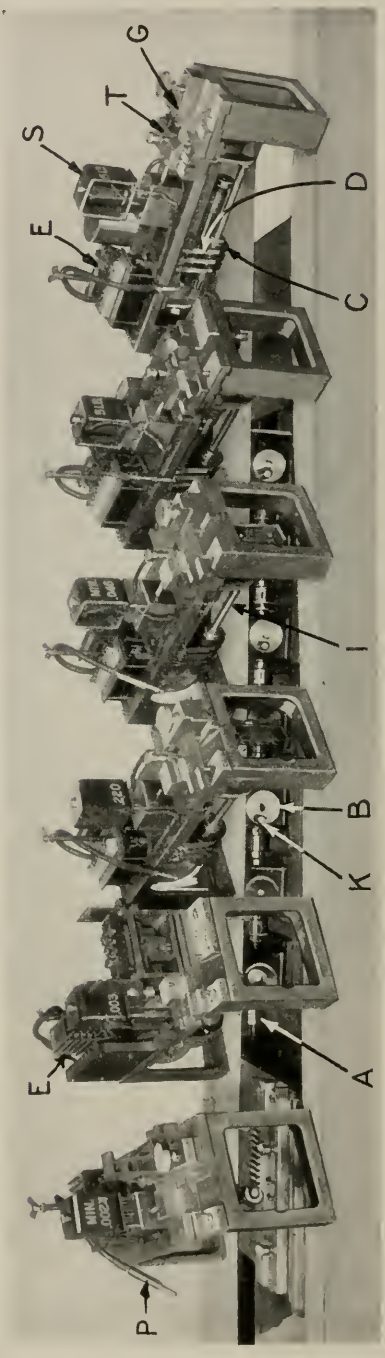


Fig. 19

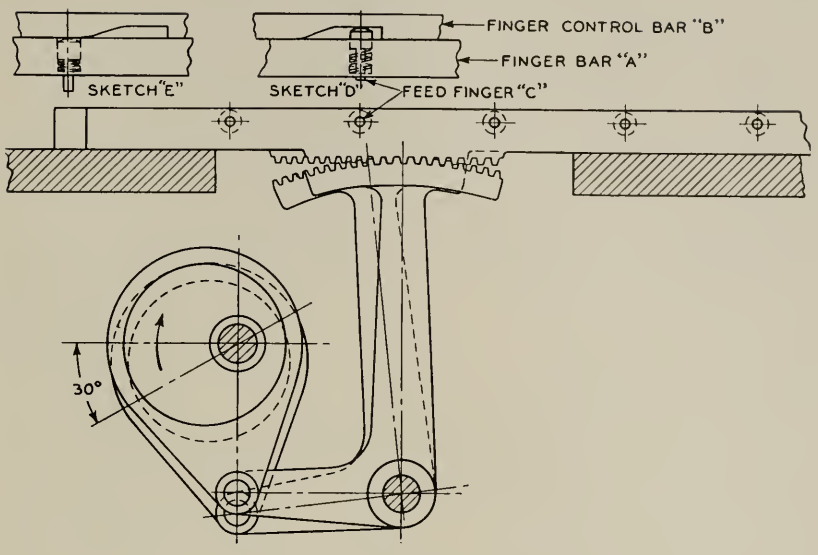


Fig. 20

The finger control bar *B* is likewise operated by a bell crank gear segment and eccentric but is set with a thirty degrees lag. The effect of this lag is illustrated in sketches *D* and *E* which show the manner in which the feed fingers *C* are withdrawn on the return stroke of the finger bar.

Economic Considerations

A comparison of hand versus machine gaging is given in Table I. In this particular case the quality of the inspection work was bettered approximately 100 per cent, while the cost was reduced 60 per cent. While this showing is rather better than the average, the tendency in most instances is in the same direction.

Like the turret machine previously described, this one was designed with the idea of making it readily adaptable for similar work on other parts. The number of units and thereby the number of operations on a machine may be varied greatly. A unit may be quickly removed for adjustment or repair and easily replaced. If the conditions warranted, spare units could be provided and an adjusted unit put in the place of a defective one in fifteen or twenty minutes.

The frame is built of welded structural steel. The machine is a complete unit with individual motor drive requiring only the attachment of the electric power supply.

The unit system for the equipment provides a wide latitude in the choice of gaging and testing fixtures to be used and the details of operating them.

TABLE I

COMPARISON OF THE ECONOMIC FACTORS ON TESTING OF PROTECTOR BLOCKS BY MACHINE METHOD AND BY MANUAL METHOD

	Machine Method	Hand Method	Remarks
Capacity	8,000,000 per year, 1 machine at 3,600 per hour	8,000,000 per year, 9 hand gages required	
Cost of Equipment	1 machine at $10,000	9 gages at $150—$1,350	Machine method requires $8,650 additional first cost, meaning an annual yearly charge, at 8%, of $670.
Cost of Labor	1½ operators at $1,920 = $2,880 per year (including loading)	6 operators at $1,920 =$11,520 per year (including loading)	Machine method gives a saving of $8,640 per year in labor costs.
Cost of Power	$40 per year	0	Machine method costs $40 a year additional for power.
Floor Space	Machine requires 35 square feet	6 operators require 90 square feet	55 square feet saved by machine.
Maintenance	Cost of maintenance at $130 per million blocks = $1,040 per year	Cost of maintenance of 9 gages per year =$1,560	$520 saved by machine method, nearly 30%.
Accuracy — Repeating results on parts that vary from tolerance limit by .0001 in.	95% (See note below)	45% Low degree of accuracy due to (*a*) plurality of gages, (*b*) plurality of operators	Degree of accuracy doubled.

Note: This means that a master gage or parts that are .0001 in. outside the tolerance limit will be rejected, in the first case, an average of 95 times in 100 trials, and in the second case 45 times. Parts that are .0001 in. within the tolerance limits will be passed as good in the same ratios. The disposition of parts that vary more than .0001 in. either way from the tolerance limits would follow the normal probability law. The figures given do not give any indication of the very small percentage of defective parts that would be passed as good or good parts classed as defectives, as these would depend upon the relative number of defectives and the distribution of their variations from the tolerance limit, as well as the precision of the methods given above.

The development of machine gaging has been greatly aided by the development of accessory parts, such as reliable indicating gages, chromium-plated parts, sensitive but sturdy relays, vacuum tube amplifiers and photo-electric cells.

Sampling methods or percentage inspection are applicable to parts that are made under conditions that may be considered approximately uniform or, as a statistician would say, under "a constant system of causes." Piece parts made in the punch press and screw machine are good examples of this. Many other classes of operations, particularly those in which some part is manual, produce parts which are not so uniform. As the variability increases, or as the requirements for precision become more exacting, the possibilities of sampling inspection become less attractive.

In many cases the conditions and requirements are such that only detail inspection or gaging is satisfactory.

In some instances automatic machine gaging of the entire output will cost less than a sampling system in which there must be included with the direct cost of inspection the cost of some additional supervision and control.

The possibilities so far as designs are concerned seem almost unlimited, so that the question of when to apply such methods becomes purely an economical one in which the number of parts to be handled, the difficulty, unpleasantness or tirescmeness of the operation, the precision required, and the cost of suitable labor become the controlling factors.

Aside from the question of cost, it is often a matter of great satisfaction to place an objectionable hand operation on the machine and release the labor for more pleasant and useful work.

Contemporary Advances in Physics, XVI

The Classical Theory of Light, Second Part[1]

By KARL K. DARROW

MEASUREMENT of wave-lengths is the subject which we shall now consider. So entitled, the topic seems unpromising, as some dry exercise in mensuration; but in truth it is distinguished for beauty and variety, and implicated with the whole of modern physics. This is not measurement of the lengths of palpable objects, as pieces of lumber or cloth, which are laid alongside of a yardstick or clamped in the jaws of a gauge. In optics, the methods of measuring wave-lengths are the methods of proving that waves exist, therefore of testing the undulatory theory of light. One could not reasonably ask for evidence of light-waves more convincing than the concord of the values obtained for the wave-length say of sodium yellow light, by all the diverse instruments which act by causing interference or diffraction: Newton's tapering film of air between a lens and a plate, Fraunhofer's grid of iron wires, the tilted mirrors of Fresnel, Michelson's echelon, and all the many gratings and interferometers continually in use in laboratories and classrooms. Wave-lengths of X-rays are computed from the diffraction-patterns imposed on X-ray beams by intercepting crystals, and these patterns were the evidence which showed some fifteen years ago that the rays are of the nature of undulations, though it could not disprove that in some paradoxical way they are also of the nature of corpuscles. From similar diffraction-patterns imposed by crystals on electron-streams it follows that these also are partly of the nature of waves, and again the patterns have supplied the values of the wave-lengths.

Moreover, evidence for waves and values of their lengths are only part of what a grating can supply. Once we are sure that we know the wave-length of a certain kind of light, we can send it against a grating and study the diffraction-pattern with the opposite intent: analyzing not the light but the grating, and deducing the widths and the spacings of the slits, if it is an alternation of slits and stops—the spacing and the shaping of its grooves, if it is an engraving on metal or glass—the arrangement of the atoms and of the electricity within the atoms, if it is a crystal. Therefore the methods for measuring wave-lengths of X-rays are also those for exploring the structures of solids and of the atoms of which these are composed. Remember

[1] Continued from the April, 1928, issue.

also that the instruments efficient in this field are the most delicate and accurate which have ever been made for any purpose; that they may be used to measure ordinary lengths and other physical quantities with an almost unbelievable precision; that the theory of relativity sprang from an experiment performed with one, and the only known way of measuring the diameter of a star involves the use of another. Surely, if this topic is not interesting, nothing in physics is interesting.

Methods of measuring wave-length are sometimes divided into those which operate by interference and those which utilize diffraction. Though to a thorough insight the distinction is only trivial, at the outset it is convenient. In an extreme example of what is specially called "diffraction," a single train of plane parallel waves is sifted through a sieve in the form of a grid or a sequence of slits; and each element of wave-front which passes through a slit evolves and spreads thenceforward according to the law of wave-propagation. Eventually—as a rule, in the focal plane of the lens beyond the grating—a region is reached where the light from the several slits intermingles; and here occur the variations in amplitude which disclose the waves and the wave-lengths. For, as I have said earlier, the eye perceives only the amplitude of the light-waves, and not their phase; therefore, in a plane-parallel beam where the phase is perpetually changing but the amplitude is everywhere the same, the eye receives a uniform impression, with nothing wavelike in it; and to make such a beam reveal that it is undulatory, we must cause the amplitude to vary from point to point. This is what we accomplish by breaking the beam into fragments, or lacerating it with obstacles, preferably with an obstacle having a periodic structure of its own, which is a grating. But it may also be accomplished by causing two plane-parallel beams to intersect one another under proper conditions. The region where they overlap is then a region of varying amplitude—indeed, the variations are as great as one can imagine; for if the beams are equally intense, there is a succession of parallel planes of no vibration and darkness, which separate spaces where there is vibration and light. The widths of these spaces, the *fringes*, may be computed from the wave-length, and reversely the wave-lengths from the widths, very simply and without any knowledge of the law of wave-propagation beyond the familiar expression for plane waves. Therefore this method of measuring wave-lengths by causing *interference* of two parallel beams is much the easiest to grasp; but it does not differ in principle from the method involving a grating, for that acts by interference between the beams from the various slits.

The Diffraction Grating

The ideal diffraction grating of theory is a sequence of equally-wide perfectly vacant slits, separated from one another by strips absolutely opaque and equal in width to one another though not necessarily to the slits. Actual gratings seldom resemble this picture, though Fraunhofer's first—the most important instrument, I suppose, in the story of spectroscopy—was an approximation to it which he made by winding a wire around and around a pair of screws held parallel and wide apart, soldering it in place and cutting away the alternate strands. So were some of his others, composed of gold-leaf mounted on glass and scratched along parallel lines with a diamond. So-called *reflection gratings* would also conform with the picture, if they consisted of bands of perfectly smooth reflecting metal separated by absolutely non-reflecting bands; for then the result would be the same as if the light came through the reflecting strips from a virtual image of the source located behind. Practical reflection gratings are not usually very like this conception, for the entire surface of the metal block is ploughed up into roughly-shaped furrows. In fact one could scarcely define the word "grating" less generally than as a periodically-repeated obstruction, or better yet a periodically-repeated device for perturbing the free onward flow of a beam of light.

Nevertheless the theory of the ideal grating contains most of what is required for the theory of the practical appliance. The reason is, that the action of the grating upon the light can be separated into two factors, each of which produces its own separate effect, each of which may be studied apart from the other. Commonly there is a set of maxima of brilliance in the diffraction-pattern; otherwise expressed, there are certain directions in which the intensity of the diffracted light is exceptionally great. From the locations of these maxima, the wave-length of the light is calculated. Now these locations are determined by the spacing of the units—be they slits and bars, furrows and ridges on a reflecting surface, planes of atoms in a crystal, or what not—whereof the exact repetition in sequence constitutes the grating. Thus if we know that a certain grating is ruled with 1000 "lines" to the inch, we can compute the wave-length of sodium light from the positions of the maxima in its diffraction-pattern, without knowing or caring whether the rulings are slits, grooves, triangular indentations, wavy ripples, or rough-bottomed troughs. If we know that in a crystal a certain grouping of atoms repeats itself one million times in a centimetre, we can calculate the wave-length of an X-ray beam or an electron-beam from the locations of its diffraction-maxima, without knowing anything about the arrangement of atoms in the group.

The contour of the rulings of a grating does, on the other hand, affect the relative intensities of the various diffraction-maxima and the details of the distribution of intensity throughout the diffraction-pattern. If they are grooves or troughs, their profiles in cross-section have influence upon the pattern; if they are slits with bars between, the ratio of width of slit to width of bar must be taken into account. In crystals the arrangement of the atoms in the groups controls the intensity-ratios among the diffraction-maxima and conversely is deduced from observations made on these. Even the distribution of electricity in the separate atoms of a crystal may be read from the details of the diffraction. These effects however can intrude upon the measurement of wave-lengths only in the cases—comparatively rare—in which some of the diffraction-maxima are actually blotted out, so that the uninformed observer may misinterpret those remaining. Except for cases such as these, one may derive the formula for computing the wave-length by assuming any convenient form of grating; and therefore we may think about a grid of slits and bars.

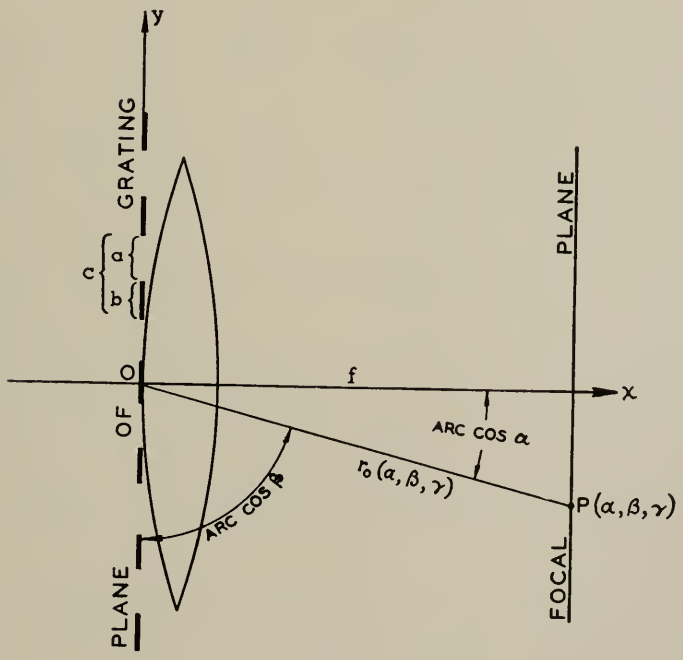


Fig. 1.

Consider then an alternation of slits of width a and bars of width b, occupying the plane $x = 0$. A beam of plane-waves, monochromatic

and of the wave-length λ, travelling along the x-direction in the positive sense, shall fall normally upon it from behind. It is our object to determine the amplitude of the waves in the region in front of the grating. Suppose for instance that we select a plane parallel to the grating and at the distance x in front of it, and derive the formula for the amplitude at any and every point (x, y, z) of this plane. This formula is the description of the theoretical diffraction-pattern in the plane in question; and the actual pattern may be observed by setting up a screen or a photographic plate in the corresponding place.

We went through this process for a single aperture in the first part of this article; and there we found that the pattern is simpler (or, at least, more calculable) the farther the plane of observation is removed from the plane of the slit, being simplest when the two are infinitely far apart. To realize this case in practice we have only to set a lens immediately before the grating. Then, the diffraction-pattern appropriate to the plane at infinity—the so-called "Fraunhofer diffraction-pattern"—is transposed into the focal plane of the lens, where we must place the photographic plate in order to record it. Naturally it is reduced in scale and augmented in intensity when it is thus transposed, and for this as well as other reasons we had better express it, not in terms of the coordinates (x, y, z) of the points in the focal plane, but in terms of the direction-cosines $(\alpha = x/r,\ \beta = y/r,\ \gamma = z/r)$ of the lines drawn to these points from the origin of coordinates.[1] Our formulæ for the diffraction-pattern in the infinitely-distant plane are in fact naturally expressed in terms of α, β and γ; and the lens may be regarded as an agency whereby that value of amplitude, which otherwise would have existed infinitely far away upon the line with direction-cosines (α, β, γ), is amplified by a constant factor and shifted inward along this line to the point where it intersects the focal plane.

We wish, then, to determine the vibration produced by a regular sequence of slits, all over the plane which is either infinitely distant or else the focal plane of the lens, according as the lens is absent or present.

Now we already have a formula for the vibration produced in that plane by any slit individually. It is the formula (93) of the first part of this article; to wit:

$$\begin{aligned} s &= \text{const. } (1 + \alpha)[C \sin (nt - mr_0) - S \cos (nt - mr_0)] \\ &= \text{const. } (1 + \alpha)\sqrt{C^2 + S^2} \sin (nt - mr_0 - \epsilon). \end{aligned} \tag{1}$$

[1] The origin should coincide both with the centre of the lens and with some point in the plane of the diffracting apertures. This is impracticable; but the error apparently does not make any trouble in practice.

Here the symbol s stands for the amplitude of the vibrating entity—whatever that may be—at the various points ("field-points") where the plane in question is intersected by the lines drawn from the origin with direction-cosines (α, β, γ); the symbols n and m for 2π times the frequency and 2π over the wave-length of the vibration, respectively; and the symbols C, S, r_0 and ϵ for various functions of (α, β, γ). In particular, C and S denote certain integrals extended over the slit, so that they involve the breadth of the slit as well as the variables (α, β, γ); this last is true of ϵ also; but r_0 denotes the distance from the origin of coordinates to the field-point, and thus involves the variables but not the breadth of the slit. As for the nature of the vibrating entity which is designated by s, I am keeping it intentionally vague. Suffice it to say that s is something of which the phase cannot be detected in any known way, but the amplitude controls the intensity of the light; the observed intensity being, according to the classical theory of light, proportional to the square of the amplitude.[2]

If therefore we were studying the diffraction-pattern of a single slit, we should be concerned only with the factor $(1 + \alpha)\sqrt{C^2 + S^2}$ in the expression for s. It would be short work to develop the expressions for C and S for a single rectangular aperture, finite or infinite in length; and having developed them, we should have solved the problem of the single slit; but in respect of our present purpose, it would be a detour. Remarkable as it may seem, the pattern of the single slit is only of secondary importance in determining that of a regular sequence of slits. When we undertake to sum up expressions such as (1) in order to compute the diffraction-pattern of such a sequence, we find the emphasis violently shifted. A new set of diffraction-maxima appear, and their positions are determined by the variation of the phase $(nt - mr_0 - \epsilon)$ from one slit to the next—in more general language, the variation which the phase undergoes in passing over one complete *period* of the grating-structure. Meanwhile the influence of the coefficients C and S, and that of the breadth of the slit which they involve, recede into the background. Not the features of the individual slit, but the interval at which one follows another, is now the dominant factor. This is the situation foreshadowed in the introductory pages.

To bring this out, let us orient the z-axis in the plane of the grating so that it runs parallel to the slits, which are of width a and are separated by bars of width b so that the period c of the grating is equal to $(a + b)$. The x-axis is to run, as heretofore, perpendicular to the surface of the grating and through the centre of the lens, so that it

[2] Or rather to the sum of the squares of the amplitudes of several quantities, any one of which separately satisfies the same equations as s.

intersects the focal plane (or the plane at infinity) at the point which is the centre of the diffraction-pattern. The y-axis is to lie in the plane of the grating perpendicular to the slits. The light comes up to the grating normally from behind, and therefore follows the x-direction. For reasons which will presently appear, it will suffice to calculate the diffraction-pattern over not the entire focal plane, but only the line where this is intersected by the xy-plane. For any field-point on this line $\gamma = 0$ and $\alpha = \sqrt{1 - \beta^2}$.

To the total vibration at any field-point, each slit now makes a contribution given by the expression (1). Numbering them in order, we may write for the contribution of the kth slit:

$$s_k = \text{const. } (1 + \alpha) A_k \sin \varphi_k, \tag{2}$$

in which A_k stands for the value of $\sqrt{C^2 + S^2}$, and φ_k for the value of $(nt - mr_0 - \epsilon)$, appropriate to the kth slit.

Now A_k has the same value for all the slits. This may be proved directly from the formulæ [3] for C and S, or indirectly by the following chain of reasoning. The function $\sqrt{C^2 + S^2}$ describes the diffraction-pattern formed by the single slit on an infinitely-distant screen, when there is no lens. Two similar slits a finite distance apart would produce two such patterns, one displaced by the same finite amount relatively to the other. But on the infinitely-distant screen the fringes and other details of the patterns are themselves infinitely broad, so that a finite displacement of one with respect to the other leaves them still practically—and, in the limit, exactly—in coincidence. This remains true when the patterns are transposed to the focal plane of the lens; those produced by a slit in one place coincide exactly with those which would be produced by an exactly similar slit lying anywhere else.[4] Therefore $\sqrt{C^2 + S^2}$ must be the same function of (α, β, γ) for every slit.

At first glance this argument seems to prove that the diffraction-pattern for the grating is merely that of the individual slit, multiplied manyfold; but that conclusion would in general be false, for we have not to add amplitudes but to compound vibrations with due regard to their relative phases. The phase φ_k which figures in equation (2) differs from slit to slit; and if these follow one another at equal intervals, φ_k changes from one to the next in equal steps.

[3] By operating on the expressions (presently to be derived) for C and S in the case of a rectangular aperture, one may show that, while each separately varies when the position of the rectangle with reference to the origin is changed, the sum of their squares remains the same. As any finite aperture may be regarded as a collection of finite or infinitesimal rectangles, the theorem is general. I am indebted to Mr. L. A. MacColl for working this out.

[4] The practical limitation to this statement would be set by the impossibility of making an ideally perfect lens of indefinitely great size.

To prove this, and to find the magnitude of these equal steps, one may proceed as follows. Omitting the lens again, consider in the grating any two consecutive slits k and $(k + 1)$, and on the very

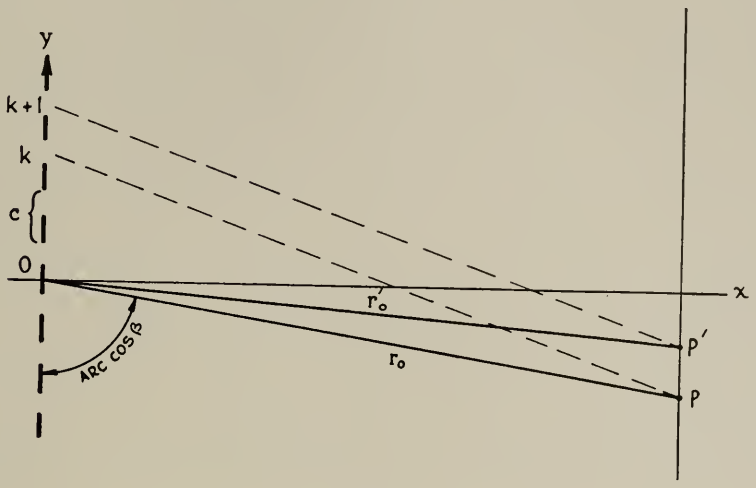


Fig. 2.

distant screen two field-points P and P' separated by the same distance c as separates corresponding points of the two slits—that is, the period of the grating. Write down successively the formulæ for the vibrations produced by k at P and by $(k + 1)$ at P'. They are, respectively:

$$A \sin (nt - mr_0 - \epsilon_k); \qquad A \sin (nt - mr_0' - \epsilon_{k+1}).$$

Since P lies in the same direction from k as P' from $(k + 1)$, these two are equal; hence:

$$\epsilon_{i+1} - \epsilon_i = m(r_0' - r_0). \tag{3}$$

Here the factor $(r_0' - r_0)$ on the right is the difference between the distances from the origin to P' and to P. In the limit when these distances become infinitely great, all the lines from the origin and the slits to P and P' become parallel and inclined to the plane of the grating by the angle of which the cosine is β; and the difference between the paths to P' and to P from the origin attains the limiting value $c\beta$. Hence in the limit:

$$\epsilon_{k+1} - \epsilon_k = mc\beta. \tag{4}$$

This is the "step" or difference in phase between the contributions of successive slits to the vibration at the field-point. The expression looks more familiar if we put θ for the angle between the normal to the

grating and the direction from the grating to the field-point, so that $\beta = \sin\theta$; then:

$$\epsilon_{k+1} - \epsilon_k = mc \sin\theta = \frac{2\pi}{\lambda} c \sin\theta. \tag{5}$$

Thus we have arrived at the conclusion—indeed almost self-evident—that the consecutive slits of the grating supply to the total vibration at the field-point contributions which are exactly equal in magnitude and follow one another at equal intervals of phase.

Our problem therefore is to sum up the series of these contributions. The process is an easy one; but we shall be able to foresee the major feature of a diffraction-spectrum without even writing down the summation. For it is evident that there must be maxima of vibration-amplitude, maxima of diffracted intensity, at the field-points or in the directions where the contributions of all of the slits agree in phase—that is to say, differ in phase by integer multiples of 2π. Counting outward from the centre of the diffraction-pattern, or normal to the grating, the first of these maxima must lie in the direction for which the "step" in phase of equation (5) is equal to 2π; hence for this "first-order maximum":

$$\sin\theta = \lambda/c. \tag{6}$$

The second lies in the direction for which the step in phase is equal to twice 2π; the third in the direction for which the step is thrice 2π; and in general there is a sequence of maxima, the general formula for the nth of which is the celebrated "plane-grating formula":

$$\sin\theta_n = n\lambda/c, \qquad n = 1, 2, 3, 4 \cdots. \tag{7}$$

The symbol n stands customarily, and in this article henceforth shall stand, for the *order* of the maximum; from now on I will write $2\pi\nu$ for the quantity which before was denoted by n.

These are the great principal maxima of the spectrum cast by a diffraction-grating. There are others between, but in practice they are inconspicuous or invisible. Those of which I have just derived the locating formula are the maxima from which wave-lengths are computed. Let it be emphasized again that the formula was derived without taking into account the ratio of slit-width to bar-width, and that it does not involve the width of the individual opening, but only the spacing between corresponding points of consecutive slits. Indeed, if one examines the deduction, it will be seen that really nothing peculiar to a slit enters into it at all. All that is preassumed is that the grating sends to the field-point a series of component vibrations, equal in amplitude and stepped off equally in phase. Such is indeed

the case when the grating is a series of windows letting light through towards the camera, separated by walls which intercept the light. Such is also the case when the grating is a series of mirrors reflecting light towards the camera, separated by windows which let it escape or by absorbing surfaces which swallow it up. Such is the case when the instrument is a surface of metal ploughed into furrows, so that the reflecting-power towards any assigned direction varies from point to point across the furrow, and varies periodically as one moves across the system of furrows. Such is the case if the waves traverse or are reflected from all points of the grating equally, but with phase-retardations which vary periodically across the grating-surface. Such is the case when the instrument is a surface containing oscillators able to vibrate in unison with the incident waves and able to radiate new waves because of their vibration, these oscillators being evenly spaced or else clustered in identical groups which themselves are evenly spaced. Such in fact is in general the case whenever the "grating" is any object with a periodic structure, the details of which are able in any manner known or unknown to perturb the passage of the waves; for anything which confuses or impedes the even onward progress of a train of waves, whether it be a vibrator which they set into oscillation or merely an inert impenetrable obstacle in their way, becomes thereby the source of a new system of undulations.

Wave-lengths of light therefore are determined by setting up in the path of the light-stream something which has a periodic structure, of which the period is known; locating the diffraction-maxima, if such there be; and using the formula (7), provided that the object is plane (another, which we shall eventually derive, is used if the object is three-dimensional and the waves travel across its structure). Location of one maximum would not as a rule suffice, for without further knowledge its order could not be identified. One must measure sufficiently many maxima to infer from the ratios of their values of $\sin \theta$ what their orders are. On the other hand, understanding of the precise mode and mechanism of the action of the elements of the grating upon the light is not required, desirable as it may be. Perhaps we do not properly understand how the atoms of a crystal scatter even X-rays; and certainly the founders of the wave-mechanics did not foresee that crystals scatter electron-waves. Yet Davisson and Germer determined the wave-lengths of these latter in 1927 with the same equation wherewith Fraunhofer in 1821 had ascertained the wave-lengths of the lines of the solar spectrum—the very equation,

$$\lambda = \frac{c}{n} \sin \theta_n,$$

taking for c the spacing between consecutive lines of atoms in the surface-layer of the nickel crystal which diffracted the electrons, where Fraunhofer had taken the spacing between the wires of the grid which was his primitive grating.

Next it is important to discover how distinct these maxima are; whether, when the amplitude of the vibration in the focal plane is plotted against θ, the peaks are broad and flattish or narrow and sharp. This too can be foretold without the labour of a complete solution of the problem. Taking any of the principal maxima—say, that of nth order, which is located at the angle $\theta_n = \text{arc sin } (n\lambda/c)$—let us inquire how near to it the amplitude will sink to zero.

Now the nth of the principal maxima is located by the condition that the phase of the contribution of every slit is $2n\pi$ in arrear of that of the slit preceding. If there are $2M$ slits altogether (it is convenient to suppose the total number to be even, though whether it is even or odd makes no appreciable difference), then at θ_n the contribution of the last slit is $(2M - 1)n \cdot 2\pi$ behind that of the first. Estimate now the vibration at the point in the focal plane —call its direction-angle $(\theta_n + \Delta)$—where the contribution of the last slit is $(2M - 1)(n + 1/2M)2\pi$ behind that of the first. It is readily shown that here the component vibration due to the last or $2M$th slit is exactly equal in magnitude and opposite in phase to that which is produced by the Mth of the slits; and in the same way every ruling of one-half of the grating may be paired off with the corresponding ruling of the other half, their effects destroying each other pair by pair. At the angle $(\theta_n + \Delta)$, therefore, there is darkness; and likewise at the angle $(\theta_n - \Delta')$, where in the focal plane or the infinitely distant plane the contributions from the first and the last slit arrive with a phase-difference of $(2M - 1)\left(n - \frac{1}{2M}\right)2\pi$.

The entire peak culminating in the nth principal maximum is consequently bounded by the directions $(\theta_n - \Delta')$ and $(\theta_n + \Delta)$; and it is easy to see that the greater the number of rulings (the spacing being supposed to remain the same) the narrower and sharper is the peak, and the more accurately can the location of its summit and therefore the wave-length be determined. Its breadth, in fact, varies inversely as the number of rulings or "lines." This is shown by writing down the formulæ for the angles corresponding to the minima which bound it. We have:

$$\sin (\theta_n + \Delta) = \left(n + \frac{1}{2M}\right)\frac{\lambda}{c}; \qquad \sin (\theta_n - \Delta') = \left(n - \frac{1}{2M}\right)\frac{\lambda}{c},$$

whence, approximately,

$$\Delta = \frac{\lambda}{2Mc}\frac{1}{\cos\theta_n} = \frac{1}{2Mn}\tan\theta_n, \tag{8}$$

so that the narrowness of the peak, as one might say, is proportional to the order thereof as well as to the total number of lines in the grating.[5] If the grating were infinitely wide, an unlimited sequence of perfectly-evenly-spaced identical units, the peaks would be infinitely narrow; light of a definite wave-length would be diffracted only in certain perfectly definite discrete directions.

Alikeness, closeness, and multitude of rulings are therefore the desiderata of a grating; alikeness, because without it the first condition for the formation of sharp diffraction maxima would be lacking—closeness, so that the maxima of lower orders, the only ones sufficiently intense to be perceived, may be spread out widely enough for convenience of observation—multitude, so that the diffracted beams shall be narrow and sharp, easy to set upon and easy to discriminate from one another. The degree of closeness which is required depends upon the spectral range which is to be explored. Ordinary "optical" gratings ruled with a diamond on metal or on glass are acceptable throughout the visible spectrum and the range to which the title "ultra-violet" is commonly restricted, extending from the visible down to wave-lengths of the order of one hundred Angstroms. They are however too fine for the remoter infra-red, for the study of which coarse lattices of wire have been used; *a fortiori* they are much too fine for Hertzian or radio waves, for which it is no exaggeration to say that a colonnade might operate as a grating; and they are commonly considered much too coarse for X-rays, though during the last two or three years several men of science have achieved the great technical feat of forcing optical gratings to measure wave-lengths which formerly were thought accessible to crystals only. Crystals are too fine for the visible spectrum, and too coarse for certain of the gamma-rays which proceed from the collapsing nuclei of atoms undergoing transmutation. Crystals with spacings of unusual width from atom-plane to atom-plane are

[5] These facts are usually expressed as statements about the "resolving power" of a grating; for if the incident light contains two not very different wave-lengths, they will form two peaks of each order not very far apart, and the possibility of distinguishing these two—of "resolving" them, to use the technical term—will depend upon the narrowness of each. If arbitrarily one says that two such peaks are just distinguishable when the summit of one falls upon the minimum adjacent to the other—in which circumstance the difference $\delta\lambda$ between their wave-lengths may readily be proved equal to the quotient of the mean of their wave-lengths, λ, by $2Mn$—then by this criterion a grating is able in its nth order to discriminate two adjacent lines of the spectrum, if their wave-lengths differ by more than that amount; and by definition the resolving power of the grating in its nth order is $\lambda/\delta\lambda = 2Mn$, the product of the number of rulings by the order.

chosen for work upon the longer X-rays, as optical gratings are ruled with lines unusually far apart for work in the near infra-red.

The ruling of good gratings is an art; and those who have practiced it with conspicuous success are fewer far than those who have attained pre-eminence in music or in painting. Amateurs, mechanics, and professors figure upon the list, the first of all being Fraunhofer, who from a glazier's apprentice evolved into the founder of spectroscopy. After his gratings of wires and of scratches in a foil of gold-leaf, he invented the method of engraving with a diamond-point upon a surface of metal or of glass (he used the latter) which is followed to this day. He met and grappled with all the difficulties which were later to beset his followers, and described them in language which now sounds strangely modern. Then, as now, it was possible to rule tens of thousands of rough grooves roughly to the inch; the trouble lay, as still it lies, in making them identical and spacing them equally.

Equality of spacing depends upon a screw, which is turned through a prearranged angle and is expected to advance through a definite distance carrying the future grating with it, whenever the diamond has completed one ruling and is waiting to begin the next. Screws as manufactured are not good enough; and anyone who aspires to be a maker of gratings must first of all procure the best available, and then devote a long and tedious time—literally years—to making it still better. Primacy in the art passed to America in the eighties of the last century, because Rowland of Johns Hopkins developed with much labour a process for removing, or at least for mitigating, the imperfections of a screw. The greater the number of rulings to be laid down side by side, the longer the portion of the screw which must be made, as nearly as humanly possible, perfect; and Michelson has testified, from unrivalled experience of many years, that the time required for the process varies as the cube of the length of the screw and width of the planned-for grating. Increase of resolving-power thus is bought at an enormous price in patience and in perseverance. A research institute is as proud of a notable grating by Rowland or Michelson or Wood, as a picture gallery of an authentic Titian or Velasquez; and the promise of a new talent is not more joyfully received, than a rumour that someone is working to perfect a yet longer screw to make a yet wider grating.

Alikeness of successive rulings depends on the endurance of the diamond. The ruling-engine is sequestered in a well-insulated room, and after the temperature has settled down to constancy is set in motion by some device worked from outside, and left to do its task in solitude. If the diamond breaks, or suffers any great change in

shape during the operation, the grating is good for nothing. This cannot be foreseen, it is not even known when it happens; to stop the process to see how things are going would be like digging up a seed to see how it is sprouting. The chance of such an accident is naturally greater, the more numerous the lines—another obstacle to the successful ruling of many-lined gratings of high resolving power.

A grating having been completed, it is removed from the engine and examined, to learn not merely whether it has been impaired by deformations of the diamond, but how—assuming it to have escaped that peril—the intensities of the various diffraction-maxima of different orders compare with one another. This is something which, as I have intimated, is controlled by the shape of the groove; this is the feature in which the individual units of the periodic structure manifest their quality. One shaping might obliterate all diffraction-maxima of even order; another might make the maxima on one side of the normal to the grating-surface stand out much more prominently than their companions on the other; still another could concentrate most of the diffracted light into one single beam. The maker of the grating cannot foresee, or can at best foresee only in part, what distribution of intensities he is going to get; for he cannot control the shape of the diamond-point, nor find it out by examination.[6] Having observed the distribution of intensities, however, he can deduce from it some facts about the shape of the grooves. This I suppose would be classified in most cases as useless knowledge; but the problem happens to be very nearly the same as that of determining the finer details of the arrangement of atoms in a crystal from the relative intensities of the various diffraction-beams which it produces when acting on an X-ray beam; and so I will devote a few paragraphs to it.

We return, then, to the grating of alternate slits and bars, to determine the influence of the ratio of slit-width to bar-width on the diffraction-pattern. Before making any calculations whatever, one striking prediction can be made directly. I have said that diffraction-maxima occur in every direction θ_n for which

$$\sin \theta_n = n\lambda/c, \qquad n = 0, 1, 2, 3, 4 \cdots,$$

because in every such direction the component vibrations arrive at the focal plane from the various slits with identical phase. But if for any of these directions the component vibrations are themselves

[6] He can control the result to a slight extent by varying the pressure with which the diamond bears upon the plate, ruling "with a light touch" or reversely; if he guesses the force just right, he may approach the condition of grooves separated by unbitten bands of smooth metal as wide as they, which resembles the theoretical case of an alternation of slits and bars of equal width.

non-existent, evidently the maxima in question are blotted out. This will happen, for example, to every maximum of even order, if bars and slits are equally wide. For, taking the direction $\theta_2(\sin\theta_2 = 2\lambda/c)$ as an instance: the contribution made to the total vibration by the upper half of each slit will be equal in magnitude and opposite in phase to that made by the lower half, and the total contribution of the slit will be zero. If in the spectrum produced by a grating the even orders are missing, or if—to say what would actually be noticed—the values of $\sin\theta$ for the present maxima stand in the ratios $1 : 3 : 5 : 7 \cdots$ instead of $1 : 2 : 3 : 4 \cdots$, the inference is that the grating has been so ruled that over half of every period the phase of the emerging (transmitted or reflected) light is constant, and over the other half no light comes forth at all; as for instance would be the case if half of every period were the unmarred surface of the metal, and the diamond had made the other half perfectly black. The reader may work out for himself what it must mean if every third, or every fourth, or every nth of the maxima is absent.

We return now to the expression (equation 2) for the contribution of a single slit or period of the grating and rewrite it, taking due account of our subsequently-gained knowledge that A_k is constant and φ_k increases by equal steps $mc \sin\theta = mc\beta$ from slit to slit:

$$s_k = \text{const.}\ (1 + \alpha)\, A \sin (nt - mr_0 - \epsilon_0 - kmc\beta). \tag{9}$$

For convenience number the slits from 0 to $N - 1$, representing by N their total number (formerly called $2M$, but now there is no reason for supposing it even), and locate the origin so that $mr_0 = \epsilon_0$. Gathering all the factors of the sine-function under a single symbol B, and writing out the expression for the summation of s_k from $k = 0$ to $k = (N - 1)$, we find for the resultant vibration in the direction θ:

$$\begin{aligned} s &= B \sum_{k=0}^{N-1} \sin (nt - kmc\beta) \\ &= B \sin nt(1 + \cos a + \cos 2a + \cdots \cos (N-1)a) \\ &\qquad - B \cos nt(\sin a + \sin 2a + \cdots \sin (N-1)a) \\ &= B\textstyle\sum_c \sin nt - B\sum_s \cos nt, \end{aligned} \tag{10}$$

in which a stands for $mc\beta$ and $\sum_c$ and $\sum_s$ for the finite series of cosines and sines which are indicated.

For the amplitude of the vibration—the only thing which matters—we then have

$$D = B\sqrt{\textstyle\sum_c{}^2 + \sum_s{}^2}. \tag{11}$$

Now, as may easily be proved [7]:

$$\sqrt{\Sigma_c^2 + \Sigma_s^2} = \sin(\tfrac{1}{2}Na) : \sin(\tfrac{1}{2}a) \tag{12}$$

so for the amplitude of the vibration in the direction θ we have:

$$D = B\frac{\sin(\tfrac{1}{2}Nmc\sin\theta)}{\sin(\tfrac{1}{2}mc\sin\theta)}. \tag{13}$$

Here we have that product of two factors which was foreshadowed in the early pages of this article—one factor (the second) depending on the periodicity of the grating, and controlling the location of the diffraction-maxima; the other depending on the structure of the individual slit or groove or atom-row, and controlling their intensity.

The second factor displays the qualities which have already been deduced by simpler means, and others. It vanishes whenever $\frac{1}{2}Na$ is an integer multiple of π, except when simultaneously $\frac{1}{2}a$ is an integer multiple of π, in which exceptional cases the great principal maxima occur. These are not the only maxima, for between any two of them there are $(N-1)$ equally-spaced minima (directions where $\frac{1}{2}Na$ is an integer multiple of π but $\frac{1}{2}a$ is not) and between these in turn there are $(N-2)$ maxima of which the locations may be found by the usual method. These so-called "secondary" maxima are however faint and inconspicuous, having, according to Wood, but 1/23 the intensity of the principal peaks, unless the grating is composed of only half-a-dozen lines or fewer.

The first factor consists essentially of that function

$$(1+\alpha)\sqrt{C^2 + S^2}$$

mentioned in equation (1) and earlier, which describes the diffraction-pattern of the single slit (or groove, or atom-row). Wherever that diffraction-pattern has a zero of intensity—the "centre of a black fringe," to use the common language—the intensity in the pattern of the grating is likewise forced to vanish. When the slit occupies half

[7] One method is based on the fact that Σ_c and $i\Sigma_s$ are respectively the real and imaginary parts of Σe^{ika}, so that

$$\Sigma_c^2 + \Sigma_s^2 = \left(\sum_{k=0}^{n-1} e^{ika}\right)\left(\sum_{0}^{n-1} e^{-ika}\right).$$

Further, by a well-known formula

$$\sum_{0}^{N-1} e^{ika} = 1 + e^{ika} + (e^{ika})^2 + \cdots (e^{ika})^{N-1}$$
$$= (1 - e^{iNa})/(1 - e^{ia})$$

and there is a corresponding expression for e^{-ika}, multiplying the two of which together and taking the square root one arrives directly at the stated result.

the width of the period (slit plus bar) of the grating, the first of its black fringes falls square upon the second-order principal maximum of the grating spectrum, which is obliterated. This is a new way of expressing the fact already mentioned, that when the slits are as wide as the bars the diffraction-maxima of even order are absent.

More generally, the intensity at any of the principal maxima is proportional to the value of $(C^2 + S^2)$ appropriate to that direction—proportional to the intensity, in that direction, of the diffraction-pattern of the single slit; and from this we can understand how, from the relative intensities of the maxima of various orders, it is possible to deduce the breadth of the slit or something about the shape of the groove. If we had only a single slit, and could send through it light of known wave-length sufficiently intense to form a measurable diffraction-pattern, we could trace the curve representing observed relation between diffracted intensity and angle θ, and compare it with the predicted curves for various values of slit-breadth; the actual width of the slit would be the value for which the agreement was perfect. If instead we had a multitude of such slits equally spaced, the observed intensities of the diffraction-maxima would supply us, not indeed with the entire continuous curve of intensity-versus-angle for the single slit, but with as many points upon that curve as there were principal maxima within our range of observation; and these—if we had two or more—would be sufficient for the comparison with the theoretical curve for the single slit, out of which the width would be deduced. From this aspect, the function of the grating is to *enhance* the intensity, at certain discrete points, of the diffraction-pattern for the single slit. Of course, when we are interested in the breadth of the single slit or the shape of the single groove, we should prefer to observe the entire continuous diffraction-pattern produced by one alone. But it may be impossible to separate one from the rest; or if we could isolate one, it might be too small to transmit or scatter any perceptible amount of light. Such is the case with atoms.

The natural gratings which atoms form in crystals are three-dimensional, and to them the reasonings which are valid for plane gratings cannot be applied without some change; but the resemblance is very close. A beam of X-rays or electron-waves falling upon a crystal is spread out into a diffraction-pattern with strong maxima, of which the relative intensities depend upon the qualities of the individual diffracting units, the atoms or the groups of atoms which are repeated over and over again to form the crystal; while their directions depend upon the spacings between these identical groups, the periodicity of the crystal. From the directions of the principal

diffraction-beams one may determine the spacings within the crystal if one knows the wave-length of the waves, or the wave-length if one knows the spacings. From the relative intensities of the beams one may deduce the distribution of the atoms within the groups, or rather the distribution of that which scatters the waves—commonly supposed to be mobile negative electricity, when the scattered waves are light; I do not know whether anyone has yet conjectured what it is that scatters the electron-waves.

The diffraction-beams proceeding from a crystal large enough to be manageable are very sharp, for the rows of atoms are far more numerous than the lines of the largest optical grating which can be made or hoped for. However, there is a limitation on their sharpness set by something to which an artificial grating is quite indifferent—the thermal agitation of the atoms, which has the same effect as though the widths of successive periods were variable and fluctuating. This effect is naturally more pronounced, the higher the temperature of the crystal; but the measurements show—for from the breadth of the diffraction-maxima it is possible to determine the mean amplitude of the temperature-agitation, another service of the crystal grating—that even at absolute zero it would not disappear, the atoms retaining a certain minimum amount of energy of vibration which apparently can never be taken from them, so long as they remain bound together in a crystal.

A few words, before leaving the subject of gratings, about the diffraction-pattern of a multitude of gratings oriented at random.

On an earlier page I said that, in computing the diffraction-pattern of a sequence of slits, we need determine it not for the entire focal plane, but only for a single line thereof—the line for which $\gamma = 0$, which is the line of intersection of the focal plane with the plane running normal to the slits and containing the infinitely-distant point-source of the parallel waves of light. The reason can now be stated. If we work out the expression $C^2 + S^2$ for a single long and narrow rectangular slit with its long sides are parallel to the z-axis, we find that the brighter parts of the diffraction-pattern form a long narrow band (criss-crossed with dark lines) with its length parallel to the y-axis and its breadth parallel to the z-axis. If the length of the rectangle grows infinitely long, the breadth of this band shrinks to zero; we have a single line of varying brightness parallel with the y-axis, which is the diffraction-pattern of the infinite slit. If instead of a single slit we have a regular sequence, their diffraction-pattern is still concentrated upon this line; it is the pattern which has just been computed, a function of the single variable β, or y, or θ; away from the line, the

intensity is everywhere zero. Spectroscopists broaden this linear pattern in practice by using as source of light not a point, but a luminous line—an incandescent filament, for instance, or a slit backed by a flame—made parallel to the slits or rulings of the grating. Then the diffraction-pattern is spread into a band. If the light is monochromatic, one sees in the focal plane, at the positions of the principal maxima, not a sequence of brilliant points as the foregoing theory implies, but a sequence of brilliant lines—the lines of the spectrum.

Instead of these lines one will obtain circles, if one uses a point-source of light and a mosaic of gratings all lying side by side in a single plane and oriented every way. Each piece of the mosaic forms its own linear diffraction-pattern, perpendicular to the direction of its own rulings; and if the pieces are numerous enough, all of these are fused into a single circular pattern, each of the principal maxima standing forth as a brilliant ring. I am not sure whether this has been done with plane optical gratings; but the analogous method with X-rays and crystals is the familiar procedure known by the names of Debye and Scherrer and Hull, or as the "powder method." Being a case of diffraction in three dimensions, it is not entirely like my imaginary case of a mosaic of plane gratings. The resemblance however extends so far, that from the broadness of the rings one may infer the size of the tiny crystals which make up the three-dimensional mosaic, the "powder"; for the smaller these are, the fewer rows of atoms each contains, and the wider their diffraction-maxima must be. But it requires very fine grinding indeed, or the dispersion of the crystals as a colloid in solution, to make them so small that the broadening of the rings is noticeable.

What would be observed, if individual slits or apertures or atoms were dispersed completely at random over the plane or throughout space? If there were many apertures all alike and all similarly oriented, but with no regularity whatever in arrangement, the diffraction-pattern would be the same as that of any singly, though more intense. The water-droplets in misty air act thus in forming haloes. If atoms were truly spherical and could be crowded together into a dense mass without any regularity, the diffraction-pattern of the mass would be that of the individual atom, and would disclose the radial distribution of its scattering-power—whether that be negative electricity, or something else. Even if atoms are not spherical, one might expect to learn in this way the average distribution of scattering-substance over all the orientations. Experiments have been conducted for this purpose; but it is difficult to find a piece of matter in which the arrangement of the atoms is entirely irregular, that is, a

perfectly "amorphous" substance; perhaps not even liquids satisfy this requirement.

Interference

When a pair of beams of light are projected together upon a screen, it is usually observed that the illumination resulting from them jointly is the simple sum of the illuminations which each produces by itself when the other is shut off. One may easily go through life without ever once finding this rule in default. Yet by intelligent design it is possible to contrive conditions in which the rule does not prevail; and actually two rays of light directed upon the same point may counteract one another and cause total darkness, and two perfectly uniform wide beams falling together upon a surface of frosted glass may decorate it with a pattern of dark fringes separated by light, dark circles alternated with bright, black networks upon a background of color—arabesques of shadow and light, more delicately shaded than anything achievable in pigment or stained glass. The brilliant and versatile Thomas Young, he who was the first to read the Egyptian hieroglyphics upon the Rosetta stone, was also the first to discover some of these lovely phenomena; a pair of exploits, which for eminence and diversity will probably never be surpassed. It happened that the first disclosure of the phenomena which demand the wave-theory of light coincided as accurately with the advent of the nineteenth century as the first realization of the necessity of quanta came at the dawn of the twentieth; for Young discovered the interference of light in 1800.

"Interference" is a name which Young selected; he said that in the conditions of his experiments beams of light interfere with one another. For the observer this was not, on the whole, an ill-chosen word, since the visible effect of the two lights conjointly is not the mere sum of the visible effects of each separately. True, it implies that the lights destroy or diminish one another, whereas in fact they are as likely to cooperate as to conflict, two equal beams combining into one of intensity as much as fourfold that of either. This is not serious; we are all accustomed to using the word *addition* to cover subtraction; and here the analogy is very close. The so-called "interference" is simply the necessary result of adding two vibrations with due regard to their direction and their phase. This is the method which was used to calculate diffraction-patterns; and in fact a diffraction-pattern is nothing but a special case of interference-pattern—not usually a simple one, for the vibrations which must be summed are very numerous, demanding integrations and long summations. The simplest interference-pattern occurs when two plane-parallel beams

of light of equal amplitude intersect one another; and this we will now consider.

Designate by 2θ the angle at which the two beams are inclined to one another, and draw the x-axis to bisect it; then the two wave-functions are

$$s' = A \sin (nt - mx \cos \theta - my \sin \theta),$$
$$s'' = A \sin (nt - mx \cos \theta + my \sin \theta)$$

and their sum [8] is

$$s' + s'' = s = 2A \cos (my \sin \theta) \sin (nt - mx \cos \theta). \qquad (14)$$

We see immediately that this is a situation in which the wave-theory of light predicts a peculiar and characteristic variation of amplitude from point to point in space, which can be tested in detail, and of which a favorably-resulting test has evidential value; whereas in either beam separately the amplitude is constant, and nothing is observable which demonstrates that there are waves. Here, in the region where the beams overlap, the amplitude varies sinusoidally between zero and the maximum value $2A$; the distance between two consecutive loci of zero amplitude, which are planes perpendicular to the axis of y, being

$$d = \pi/m \sin \theta = \tfrac{1}{2}\lambda/\sin \theta. \qquad (15)$$

The presence of a series of equally-spaced planes of darkness, their separation varying inversely as the sine of the angle between the beams, is then to be taken as evidence that light is undulatory; and from their separation and the angle between the beams one may compute the wave-length of the light. A more thorough test, made by measuring the distribution of light-intensity between two such planes, would lead (anyway it ought to lead) to the conclusion already known, no doubt, to all the readers of this paper: that the intensity of the light varies as the square of the amplitude of the waves.

To produce this effect of interference, the two intersecting beams must have started from the same source of light, and at very nearly the same instant—that is to say, the optical paths from the source along the two beams to the region of overlapping must be the same within a few millions of wave-lengths, or a few hundreds of centimetres. By the wave-theory, this is easily understood. We must think that

[8] To add them thus implies that the quantity denoted by s is either a scalar, or a vector perpendicular to the xy-plane. Since light is not adequately described by either assumption, we must anticipate defects in the theory, more prominent the larger the angle θ. In practice θ is evidently always so small that there is no trouble from this source.

a beam of light from a flame or an arc consists of myriads of feeble beams each proceeding from a single atom. Each is divided—the methods of division are the methods of producing interference-fringes—and the separate parts are then caused to overlap. Each pair which came originally from a single atom produces a set of interference-fringes, and the fringes for all these pairs coincide in space. Each fraction of a divided beam may also interfere with a fraction of another, proceeding from another atom; but owing to the uncontrolled and uncontrollable phase-differences between the beams of a pair so formed, the fringes for these pairs do not coincide, and on the whole they efface one another. By the quantum-theory the explanation—not indeed of the fact that interference occurs only under these special conditions, but of the fact that it ever occurs at all—is not so easy. Indeed the fact commonly expressed by saying that light from a source is "coherent" with itself, has been regarded as the most difficult of all for the quantum-theory to explain.

To produce interference, then, we must divide a beam of light and cause its parts to cross each other's paths. The simplest of the devices which effect this were invented by Fresnel; a pair of prisms which turn two portions of the beam towards one another, and a pair of mirrors which reflect two portions across each other's routes. A single mirror indeed suffices; standing acoustic waves are produced thus, in Kundt's tube and otherwise, with values of the angle 2θ sometimes as great as 180°; but light-waves are so short that with so great an angle the distance between dark fringes would be too small to measure, if not indeed to perceive; and we must use the facility for expanding them which the factor $\sin\theta$ in equation (15) offers us.

The Interferometer

In the devices which I have thus far mentioned, the interference of overlapping wave-trains oblique to one another causes the formation of alternate zones of darkness and light in space; and the visible fringes are the cross-sections of these zones upon a screen set up to intersect them. There are however other instruments in which the overlapping beams are parallel to one another, the region which they occupy is not traversed by bands of light and shade, and a screen thrust across it shows uniform illumination; and yet when the eye or the camera is located in that region, fringes are produced upon the retina or on the plate by the action of the lens of either. These are not so easily understood as the earlier devices, and yet it is important to comprehend them, for the striking applications of interference have been made by means of such as these. Among them are the interferometer of Fabry and Pérot, and that of Michelson.

Imagine, at the outset, a pair of perfectly plane and parallel mirrors, onto which wave-trains of extended plane wave-fronts are falling from every direction. The mirrors must of course be semi-transparent, so that part of the light which falls first upon one—say, the upper—is reflected from it at once, and part goes on to meet and be reflected by the lower. Thus (as Fig. 3 shows more clearly than words) the

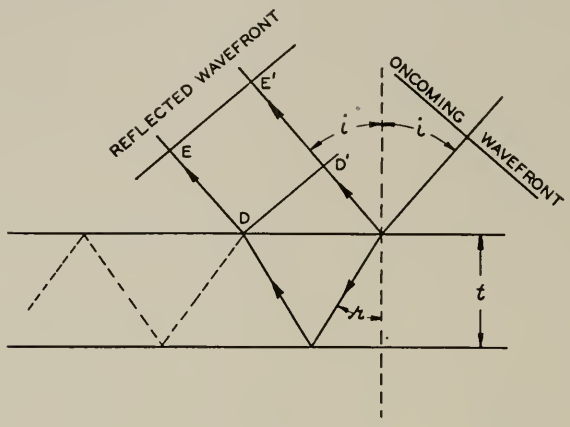


Fig. 3.

mirrors form out of each incident wave-train a first and a second reflected beam, which travel back through the space above the mirrors in the same direction, making according to the law of reflection the same angle i with the normal as the incident wave-train did. In truth there are not merely two reflected beams derived from each incident one, but an infinity thereof, owing to the multiple reflections which are indicated in the sketch. We need not however (as I shall presently show) take account of more than two; by combining the second reflected beam with the first we can predict the most important features of the interference.

It is necessary to be somewhat more precise about the nature of the mirrors. As good an example as any to begin with is that of the "thin plate"—a slab of some transparent substance, glass for instance, embedded in a transparent medium which I will take to be empty space. The mirrors, then, are the upper and lower sides of the plate. Denote by μ the ratio of the speeds of light in the environing medium and in the substance of the plate, by i the angle of incidence of any wave-train and by r the angle of refraction of its transmitted part; then as heretofore we have

$$\sin i = \mu \sin r. \tag{16}$$

The ratio A_2/A_1 of the amplitudes of the first and second reflected beams, and in general the ratios A_n/A_1 of the amplitudes of any of the reflected beams and the first, are determined altogether by μ and i. An important consequence of this will presently appear. One can however alter these ratios, e.g., by half-silvering the sides of the plate; and the formulæ which I am about to quote may be applied to the case of two half-silvered mirrors facing each other in air, by setting $\mu = 1$.

Isolate then in mind a single incident wave-train. Denote by i its angle of incidence upon the upper surface of the glass; by t the thickness of the plate. A wave-front of the oncoming wave-train is divided into two. During the time while the part which entered the glass is advancing to the lower side, being reflected, returning to and re-emerging from the upper side, the part which was first reflected goes on to the level EE' of Fig. 3. The emerging wave coincides with the first-reflected part of a new wave-front which was following along after the old one at the interval $E'D'$. In general, there is a phase-difference φ between these two. The condition for interference is ideally satisfied; two wave-fronts coincide. The amplitude A of the resultant light travelling away from the mirrors is obtained by the prior formula from the amplitudes A_1 and A_2 of the first and second reflected beams and their phase-difference φ:

$$A^2 = A_1^2 + A_2^2 + 2A_1A_2 \cos \varphi. \tag{17}$$

It is now our affair to deduce the value of this difference in phase. This is a simple task, for the angle φ, expressed in radians or in degrees, is 2π or 360 times the number of wave-lengths intervening between the wave-front DD' and the wave-front EE' which, so to speak, has gone on ahead leaving part of itself behind to combine with DD'. We have therefore only to multiply $2\pi/\lambda$ or $360/\lambda$ into the distance $D'E'$; which distance is found [9] by very easy manipulations of trigonometry, aided by the equation (16), to be $2\mu t \cos r$; so that for the phase-difference in radians we have

$$\varphi = \frac{2\pi}{\lambda} 2\mu t \cos r + \varphi_0. \tag{18}$$

The constant φ_0 is inserted to leave room for the possibility that abrupt changes of phase may occur at the instant of reflection, unequal for the two reflecting surfaces. Experience shows that in this case of

[9] For the distance $D'E'$ is the difference between BD' and BE'. The latter is evidently $BD \sin i$, which is $2t \tan r \sin i$, which is $2\mu t \tan r \sin r$. The former is the distance cT traversed by light in vacuo during the time T while the beam which entered the glass is traversing its zigzag path BCD; this time is $(\mu/c)BCD$, which is $2(\mu/c)t \sec r$.

the plate the value π must be assigned to φ_0; as though the phase were unaltered at the first reflection, but reversed at the second. This is however a point of minor importance.

The equation (18) is one of the most important in optics; we shall encounter it repeatedly, even so far along as in the X-ray range.

By comparing (17) and (18) one sees that, if wave-trains of equal intensity fall upon a glass plate from all directions, those which depart in the various directions are not equally intense; their brightnesses depend upon their angle of reflection which is their angle of incidence, i. However they are not separated in space, and hence there are no bands of light and darkness. But if a lens be placed in the region above the plate, it will direct each of the beams to a separate point in its focal plane; and since the illumination at the point where a wave-train converges is proportional to the intensity of that wave-train, there will be fringes in that focal plane. If the lens is set with its axis normal to the planes of the mirrors, as when one looks straight at them with the eye, then the points where the wave-trains reflected at any angle i converge lie all upon a circle, its radius depending on i. The fringes are therefore circular; looking vertically down upon a thin plate, or photographing it with a camera pointed directly towards it, one sees or registers a system of concentric rings, their centre wherever the perpendicular dropped from the lens reaches the plate. These are said to be localized at infinity; but the term is not a good one, for the fringes are not at infinity; they are on the retina or on the camera film, formed by the lens in the focal plane thereof.

The values of i for which the amplitudes of the reflected beams are least or greatest may be determined by differentiating (17). If we may neglect the variation of A_2/A_1 with i (as usually but not always we may), the result is the expected one: least intensity and blackest point of a fringe corresponds to a phase-difference of π or an odd-integer-multiple thereof, greatest intensity and brightest point of a fringe occurs with a phase-difference of zero or any even-integer multiple of π. Thus one may compute the actual size of the circular rings to be produced when a given lens sorts out the rays reflected by a given plate, and test the predictions by experiment; or rather, to say what is really done, one may determine the wave-length of the light by measuring the diameters of the rings and comparing them with the formulæ. But this is not a customary way of determining wave-lengths.

At this point it is expedient to remark that the size of the fringes is not affected by the presence of those third, fourth, fifth and indefinitely many reflected beams which I excluded from the computation. For the phase-difference between each of these and the one reflected once-

less-often is given by the first term on the right in equation (18), and hence is greater by 180° than that between the second and the first; and when i is so chosen that the second and the first are in opposite phase, then all the beams of higher order are in the same phase as the second and reinforce it, the reinforcement being just so great—in the case of the transparent plate—that the resultant of all these beams together is of the same amplitude as the first reflected beam. Therefore the minima, the centres of the dark fringes, are not shifted by the extra beams, but are rendered absolutely black.

The width of the fringes is greater, the thinner the plate—other things, naturally, being left unchanged. This is the reason why one needs quite a thin lamina of glass to see them well, and cannot get them at all with a windowpane. If the thickness of the plate is changing continually, one sees them narrowing or widening; one sees also a phenomenon much more striking, the generation of rings one after the other out of the centre of the fringe-system—if the plate is growing thicker; the reverse, if it is shrinking. Glass plates capable of shrinking or thickening at will are not as yet available; but at times the former case is realized by a soap-bubble on the verge of dissolution. Where the soap-film is about to give way, the fringes rapidly dwindle and are swallowed up into the central spot of the interference-system, which alternately turns dark and light, and finally goes black just at the instant before the bubble bursts. From this final blackness we infer that the value π must be assigned to the constant φ_0 of equation (3). Reflection from water to air and reflection from air to water are accompanied by phase-changes differing from each other by π, since the beams of light formed by two such reflections destroy each other when the reflecting surfaces are immeasurably close. But the bubble is too tender an instrument for practical use. The plate of variable thickness must be a slab of empty space between movable solid mirrors.*

The interferometer of Fabry and Perot is precisely such a thing: two half-silvered plates of glass facing each other across a narrow gap. The like-named instrument of Michelson is the same in principle; but by ingenious use of a third reflector, the two essential mirrors are transposed far apart—a most valuable feature, as we shall see. The incident wave-trains (coming from below, in Fig. 4) are divided by a semi-transparent mirror C inclined at about 45° to their path, and the fractions are sent off at right angles to one another, along the two

* Or else Lummer's device of a pair of wedges so proportioned, that a face of one may slide along a face of the other, while the two sides not in contact remain parallel to one another.

"arms" of the machine, at the ends of which they meet fully-reflecting surfaces A and B set normal to their courses. Reflected straight back along the arms, they are once more divided by the semi-transparent

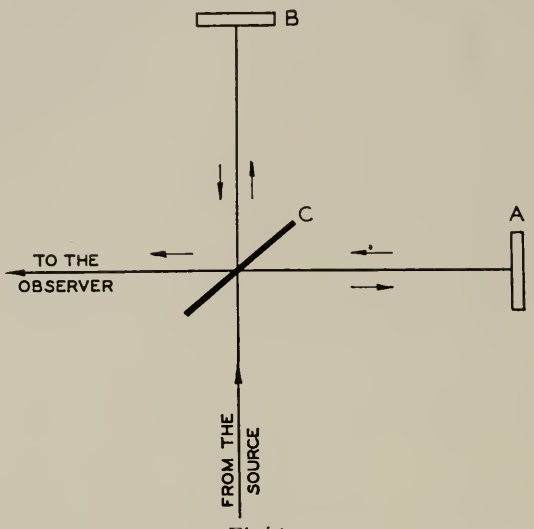


Fig. 4.

mirror, and the fractions which go towards the observer (left, in Fig. 4) are those which interfere. This seems complicated; but in essence it is not so. Everything happens as though the mirrors A and B were at the end of the same arm. Looking from the left through C, one perceives the mirror A and the virtual image in C of the mirror B; and this comes to the same as though C could be removed, and the horizontal arm of the machine swung into coincidence with the vertical arm in spite of the well-known prohibition against two objects occupying the same place at the same time. When the plane of A is accurately normal to the plane of B and the semi-transparent surface C is properly oblique, the virtual image of B is accurately parallel to A; and the observer sees circular fringes, which one by one emerge from a central spot or shrink down into it as either mirror is displaced along its arm.

This clever device of combining one real mirror with the virtual image of another, to form a thin plate of which one surface is impalpable, is of enormous value. One can make the virtual reflector go right through the real one, the fringes being swallowed up into the centre one after another, and then reborn in due order after the whole field of view goes black at the instant of coincidence. By moving one

of the reflectors through a chosen distance and counting the fringes which are born or consumed during the motion, one may evaluate the distance in terms of the wave-length of the light, or the wave-length in terms of the chosen distance, according as the one or the other is independently known. For, returning to equation (18) and putting $\mu = 1$ (since the two surfaces of the "thin plate" are separated only by air) and $r = 0$ (since the light is normally incident), we see that φ is changed by 2π when the spacing between the reflectors is changed by $\frac{1}{2}\lambda$; but when φ is changed by 2π, a single ring is added to or subtracted from the system of annular fringes; hence the total number of rings created or destroyed during the motion of the mirror is equal to twice the number of wave-lengths comprised in the distance which it traverses. In this way Michelson counted, as the first step in his determination of the length of the standard metre, the waves of the red cadmium line covering the distance between the two ends of an "intermediate standard" or *étalon*, about half a millimetre long.

In practice the real and the virtual reflector are frequently not quite parallel with one another; and this is sometimes a convenience. If they intersect (another of the things which are not possible with a pair of real reflectors) and the lens of eye or camera is located vertically above the line of intersection, this line stands forth embodied as a fine straight black fringe—the *central fringe*—companioned on either side by a multitude of others.[10] If now either of the reflectors be set

[10] Imagine a pair of mirrors inclined to one another at a very small angle φ. Establish a coordinate-frame such that the z-axis is the line of intersection of the two mirrors, the x-axis lies in the plane of either. Locate the lens of the eye or the camera at any point, say P; drop the perpendicular from P to the zx-plane at P_0; let R stand for its length, and t_0 for the distance between the mirrors at its foot. Consider the pair of reflected wave-trains arriving at P from any direction, making an angle i with the aforesaid perpendicular. Denote by α and β the projections of i upon the xy-plane and the yz-plane respectively. Assume all these angles to be small. The pair of reflected wave-trains arise from the reflection, at first and second mirrors respectively, of a single primary wave-train which fell at the same angle i upon the mirrors at a point where the distance t between them is equal to $(t_0 + R\cdot\tan\alpha\cdot\tan\varphi)$; or, to first approximation, $t = t_0 + R\alpha\varphi$. The phase-difference between them, by equation (3), is equal to $(2\pi/\lambda)2t\cdot\cos i$. To first approximation we have

$$\cos i = 1 - \tfrac{1}{2}i^2 = 1 - \tfrac{1}{2}(\alpha^2 + \beta^2).$$

Hence to this approximation we have for the phase-difference:

$$\delta = \frac{4\pi}{\lambda}(t_0 + R\alpha\varphi)(1 - \tfrac{1}{2}\overline{\alpha^2 + \beta^2}).$$

Rearranging, and dropping the terms in $\alpha^2\beta$ and $\alpha\beta^2$, we get

$$\alpha^2 + \beta^2 - 2(R\varphi/t_0)\alpha - 2 + (\lambda\delta/4\pi)(2/t_0) = 0.$$

The loci of constant phase-difference, and hence the fringes, are circles centred upon the line inclined at angle $\alpha_0 = R\varphi/t_0$ to the perpendicular dropped from the lens to the mirrors. If this perpendicular meets the mirrors along their line of intersection, as assumed in the text, we have $t_0 = 0$; the centre of the circles is infinitely remote, the fringes are straight. If then either the lens or the line of intersection is shifted sidewise, the fringes march sidewise and acquire a curvature.—It should be realized that if the wave-fronts are wide and the mirrors not perfectly parallel, there are fluctuations of intensity along each wave-front; it may be necessary to narrow the aperture of the lens in order to avoid confusion due to these.

into motion, the fringes march off sidewise, growing more curved as they go; and the number passing any fixed marker set up in the field of vision is the double of the number of wave-lengths comprised in the distance through which the mirror moves.

Consider now for a moment what must be observed, if wave-trains of many wave-lengths fall upon the mirrors, instead of pure monochromatic light. Each kind of light produces its own pattern of fringes; but since the breadth of a fringe depends upon the wave-length, those of one kind cannot fall perfectly—light upon light and shade upon shade—upon those of another; and in most parts of the field of view, if not in all, the various patterns must blot one another out. Yet there is one exception; returning to equation (18) one sees that for any value of r the phase-difference φ must vary with λ, *unless* $t = 0$—in which exceptional case $\varphi = \varphi_0 = 180°$ whatever the wave-length.[11] When the real mirror and the virtual one coincide perfectly, the field of view is black, however many wave-lengths are contained in the incident light; and when the real mirror and the virtual one intersect, the line of intersection is marked with a black fringe. Moreover, in the neighborhood of the central fringe there is a brilliant display of colors. Words are too feeble to describe them, but there would be no great scientific advantage to be gained from a description; for the tint observed in any particular direction is not a pure prismatic hue, but results from mixture of the wave-lengths not completely extinguished by interference in that direction, and depends therefore upon the physiology of the eye. What interests us as physicists is the service of the central black fringe and the companioning glory of colors in marking the point and moment when the real and the virtual mirrors intersect or coincide. This service was essential in the measuring of the standard metre, a process which we will now examine.

The interferometer used in the process is sketched as seen from above in Fig. 5. The mirrors M and M′ are at the two extremities of the intermediate standard; they are made parallel with the greatest possible exactness, and the further is elevated above the level of the nearer. As for the other mirrors, D is the movable "reference" reflector; the purpose of N will appear directly; N′ may be ignored.

The instrument is now so adjusted that D is strictly parallel to the virtual image of N, and intersects that of M at a small angle. Red cadmium light being used, a part of the field of view is occupied by the circular fringes due to the cooperation of D and N and part by the

[11] Also φ is independent of λ if $\cos r = 0$, a case which may occur if we have a stratum of air between glass plates, and choose an angle of incidence near the angle of total reflection.

straight fringes due to that of D and M. Among these latter the central fringe marking the line of intersection of D with the virtual image of M could hardly be identified; but when the white light is turned on, it stands forth unique. It is made to coincide with some

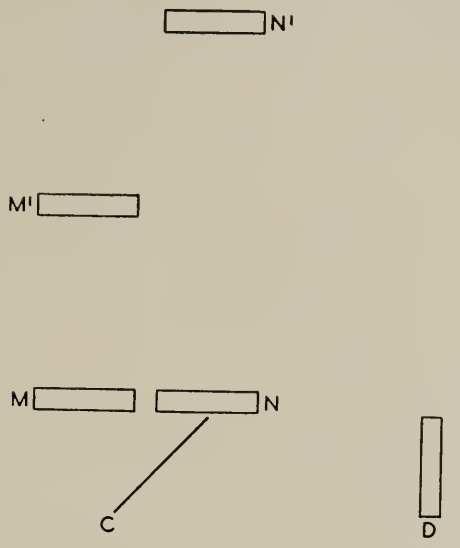


Fig. 5.

fiducial mark in the field of view, and the measurement commences. The red light being restored, the observer transfers his attention to the circular fringes formed by N and D, and counts them as they pass while he moves D along towards the virtual image of M'. Making occasional tests with the white light, he eventually notes the advent of the blaze of colours and the central black fringe which indicate the intersection of this image with D. When the black stripe coincides with the fiducial mark, the reflector D has moved through the length of the standard MM', and this length has been measured by the count of the circular fringes which meanwhile were passing by in the other part of the field of view. The number of waves is, of course, not as a rule an integer; the fractional excess may be estimated quite accurately in various ways.

In the actual work, this shortest standard was about 0.4 mm. long, and comprised somewhat over 606 waves of red cadmium light. The counting of the 1212 fringes corresponding to its length was the only counting required; the rest of the process consisted of nine stages, the first of which was the type for all the others except the last. This second step was the comparison of the shortest standard with a

second, made as accurately as possible to be twice as long; or rather, the shortest standard had been constructed with the deliberate aim of making it as nearly as possible just half so long as the second shortest. It was the office of the interferometer to determine how nearly this ideal had been attained; which was fulfilled by means of coincidences, detected as before through the coloured fringes.

Returning once more to Fig. 5, let M and M′ stand for the mirrors of the shortest standard, N and N′ for those of the second shortest. N and N′ are on a higher level than M and M′, and the field of view may therefore be divided into quadrants, in which appear the fringes—if and when there are any—formed by the collaboration of D with the virtual images of M and M′ and N and N′, respectively. The observer then goes through the following routine: (1) D is made to intersect the images of M and N; (2) D is drawn back till it intersects the image of M′; (3) the shortest standard carrying M and M′ is drawn back till the image of M again intersects D; (4) D is drawn back until it intersects the image of M′. The witness of intersection is always the central black fringe, appearing in the required quadrant or quadrants. Now if the distance NN′ is exactly twice the distance MM′, then when the observer completes the four stages of the routine D will intersect not only the image of M′ but that of N′; at the end of stage 4 the central fringe will appear in each of the upper quadrants, just above the point where at the beginning of stage 1 it appeared in each of the lower quadrants. If NN′ departs by a fraction of a wave from the doubled value of MM′, the central stripe in one of the upper quadrants will lag by a little behind that in the other. From the lag expressed in terms of the fringe-width, one may compute the difference between the length of the second standard and the doubled length of the first, and so obtain the number of waves comprised in the second. Now the second standard is made as nearly as possible of one-half the length of the third, the third of the fourth, and so on up to the ninth, which is made as nearly as possible one-tenth of the length of the standard metre. From each to the next the comparison is made in the same way. To show the remarkable reliability of Michelson's results I quote the three values which he obtained by three independent measurements of twice the number of waves of red cadmium light comprised in the length of the ninth standard:

310678.48, 310678.65, 310678.68.

After this point it remained to compare the ninth standard with a metre-rod, and the metre-rod with The Standard Metre. Returning for the last time to Fig. 5, we take M and M′ as the mirrors at the

ends of the ninth standard. The steps are: (1) make M coincide with the end of the metre-rod, and D intersect the virtual image of M; (2) draw D back to intersect with M′; (3) draw MM′ back until M coincides with D; (4) draw D back to intersect with M′—and so forth ten times altogether, until for the last time D intersects M′, and M′ is very near the far end of the metre-rod. The discrepancy is again a fraction of a wave-length.

The result in which all this labour culminated was: **1,553,163.5** wave-lengths of red cadmium light are comprised in the length of the standard metre.

Such was the process of enumerating the millions of light-waves required to make up the length of that standard chosen for the measurements of common life, and so very ill-adapted to magnitudes of the scale of those in light—the distance between two scratches on the bar of platiniridium alloy, conserved in the vault at Breteuil with the care lavished upon a sacred relic. The achievement of Michelson was the bridging of a gap, or let me say a work of translation. Nearly all measurements of wave-lengths to this day are determinations of the ratio of one wave-length to another, as practically all measurements of objects an inch, a metre or a mile long are determinations of the ratios which they bear to metre-sticks. In dealing with tangible objects we use the language of the metre; in dealing with light-waves, we use effectively the language of another scale of measurement. Michelson was the first to make a supremely accurate translation from one to the other of these languages, so making it possible to express a measurement of either realm in the scale familiar for the other. Whether in addition he may be said to have replaced a standard essentially impermanent and transitory by one which in the nature of things is everywhere the same and forever immutable, is a question very likely never to be answered.

Harmonic Production in Ferromagnetic Materials at Low Frequencies and Low Flux Densities

By EUGENE PETERSON

Synopsis: When a multi-channel communication circuit includes a non-linear element such as a ferromagnetic core coil, distortion of the wave form impressed upon the circuit is produced. In terms of the single frequency components, this distortion is manifested in the appearance of new components. This distortion may give rise to the reduction of quality in any channel, and it may also introduce crosstalk and interference, which consists of new frequencies not present in the impressed wave of any channel under consideration, produced by independent channels. In view of the recent increased use of multi-channel systems, it has become necessary to investigate the effects of this type of distortion, to determine the dependence of this distortion upon the properties of the magnetic materials constituting the cores of inductance coils and transformers, as well as upon the circuit impedances, and to determine those constants of core materials which are significant in the distorting process.

The behavior of magnetic materials to complex waves of magnetizing force is ordinarily a highly involved process, so that a direct correlation between distortion and some of the easily measured constants of materials is a matter of some difficulty. It has been established experimentally, as a confirmation of theoretical speculations, that the third harmonic e.m.f. generated by a sinusoidal wave of magnetizing force may serve as an index of the distortion with a complex wave of magnetizing force. This relation is valid for low flux densities and for frequencies at which the screening effect of eddy currents is not important. The paper is therefore devoted to an investigation of the third harmonic production in its dependence upon the properties of hysteresis loops. These loop constants in turn are shown to be deducible from AC bridge measurements on a coil of known dimensions having a core of the magnetic material under investigation. The loop constants for a few materials are included in the text. An analogy exists between the treatments of hysteresis loop and of three-element vacuum tube characteristics which enables us to compare simplifying relations introduced by Rayleigh and by H. J. van der Bijl in the two cases.

The theoretical deductions are found to be in general agreement with experiment, and are applied to a number of cases of practical interest. These include the effects of air gaps and dilution, and the choice of core material in third harmonic production by inductance coils and transformers. Finally, the amount of third harmonic current flowing out of long lines is deduced with both lumped and continuous loading.

Part 1. Hysteresis Loops and their Mathematical Representation

NEW and improved systems of multi-channel communication which have come into use during the past few years have imposed rigorous requirements on the circuit elements constituting the communicating link, and have made it necessary to investigate the degree of distortion which arises from the use of ferromagnetic apparatus. The distortion introduced may have two general effects: distortion of the signal in any one channel, which is usually the minor effect, and production of crosstalk and interference between the various

channels to which the non-linear element is common. Such elements are, in general, ferromagnetic core coils or transformers. The distortion introduced by the non-linear relation between flux density and magnetizing force is therefore of fundamental importance in the design of iron core coils and transformers which carry simultaneously a number of communication channels.

The use of iron core coils and transformers in communication work is confined to comparatively low flux densities in contrast to the ordinary practice in power work, where operation usually occurs above the knee of the normal magnetization curve at a value of the order of a thousand times greater than that used in communication work. There are two main reasons for this restriction: losses are reduced, and the relation between flux density and magnetizing force approaches linearity so that distortion is minimized.

Under actual operating conditions in which the more important crosstalk effects arise, we have a complex wave of magnetizing force acting on the magnetic core. Non-linearity in the magnetic circuit gives rise to new frequencies which normally [1] are related to those impressed upon the circuit, being sums and differences of integral multiples of the originally impressed frequencies. These modulation products are all of odd order [2] when the core is unpolarized. On purely theoretical grounds we would expect to find relations between the amplitudes of the different frequencies resulting from any one order of modulation. To take the third order modulation products of two impressed frequencies (f_1, f_2) as an example, we would expect the amplitudes of the harmonics $3f_1$ and $3f_2$ to be related to the amplitudes of the other third order products: $2f_1 \pm f_2$, $2f_2 \pm f_1$. This has been confirmed by direct experimental test, so that we are enabled to use the generated third harmonic voltage as an index of the generated voltages corresponding to the other third order products. Accordingly, we shall deal in the following with the third harmonic produced by a sinusoidal wave of magnetizing force, and so avoid a more involved analysis.

The fundamental relation in the operation of ferromagnetic apparatus is of course the relation between the flux density B and the magnetizing force H. In contrast to the usual behavior of circuit elements, the relation between the two fundamental quantities—the independent and the dependent variables, H and B in this case—is a function not only of the value of the independent variable, but is also

[1] This is true in the low flux density region. At high densities and with highly reactive circuits as in magnetic modulators, other frequencies are sometimes found which correspond to natural oscillations of the coil and circuit.

[2] *Bell System Technical Journal*, Jan. 1928, pp. 110, 111.

a function of its previous history as manifested in the phenomenon of hysteresis. This complicates matters to an extent far greater than is the case with other circuit elements, and some analysis has been carried out in which the hysteresis loop has been replaced by the normal magnetization curve, or by some such single valued relation between the variables. For some purposes this convenient simplification—it cannot be called a close approximation for our present purposes—is satisfactory, while for others it does not begin to tell the story. Inasmuch as our aim here is to deal with the actual phenomena involved rather than to arrive at some arbitrary procedure for representing the facts, we shall in the following base considerations upon the hysteresis loop.

Fig. 1 will serve to illustrate the effect of previous history upon flux density. The main loop there, which extends between the two

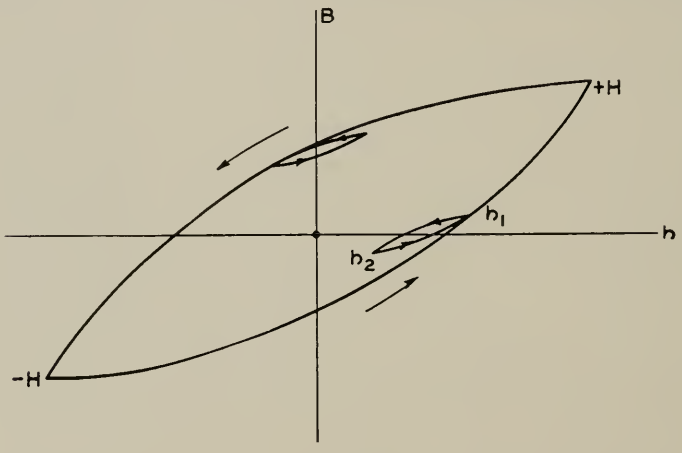


FIG. 1

limits of magnetizing force $-H$ and H, is obtained when the magnetizing force is varied cyclically between those two values in such a way that the magnetizing force has but one maximum and one minimum per cycle. In that case the B–H loop is traversed in the direction shown by the arrow. It is independent of frequency when the eddy-current losses in the iron are small, as we shall suppose them to be, and it is independent of the wave form of the magnetizing force so long as that wave form satisfies the condition we have laid down above. A sinusoidal magnetizing force, for example, satisfies that condition. When the magnetizing force contains components of such magnitude and phase that the wave form has multiple maxima or minima, the simple B–H loop no longer suffices to represent that relation, but auxiliary loops shown in the figure are involved.

Thus suppose a subsidiary maximum to exist at h_1; as the magnetizing force decreases, the flux density no longer follows the main loop but branches off on a subsidiary loop as indicated by the arrows. When the subsidiary minimum at h_2 is reached a new branch is started which completes the subsidiary loop, and which brings the magnetizing force back to the main loop at the point from which it originally diverged, and the main loop is thereafter followed until another maximum (or minimum as the case may be) of the magnetizing force wave is reached. For simplicity in the following we are going to deal solely with a sinusoidal wave of magnetizing force so that subsidiary loops are never called into play. With this understood, the relation between B and H is described by the simple loops of Fig. 2 over a

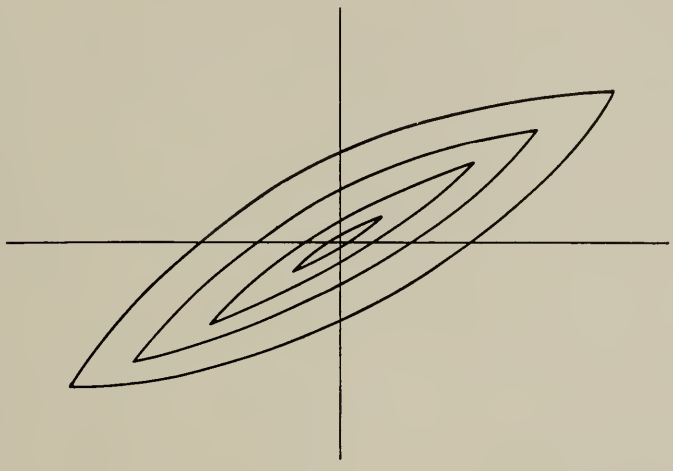


FIG. 2

certain range of magnetizing force, each loop being defined by a particular value of maximum magnetizing force. Each loop may be considered as constituted by two branches which join at the maximum field of the loop.

It is clear, therefore, that with a periodic magnetizing force having but one maximum and one minimum per cycle, the flux density depends upon three properties of the magnetizing force—the maximum value, the instantaneous value, and the sign of dh/dt, being located on the lower branch when dh/dt is positive, and on the upper when it is negative. Now it is of course evident that when a definite loop form is available, a numerical solution by graphical or step-by-step methods may be had. It is further evident that, in the case of a definite impressed magnetizing force, the B–H loop may be broken up into its

harmonic components, as was done by S. P. Thompson.[3] Solutions in this form possess all the advantages and disadvantages of numerical ones in which any change of conditions leads to a new problem; the solution desired is a general one which will describe the phenomena in terms of coefficients characteristic of the magnetic material, and this type of solution is the subject of analysis in the following pages.

Perhaps the least difficult method of arriving at an analytical solution without making any assumptions as to the form of the loops is the following. A power series for each branch of any loop is formulated, and the two resultant equations are combined in a trigonometric series. In this way the solutions, each one valid over but half the cycle, are combined to represent the relation of B to H over the entire cycle.

The flux density on each branch of the loop may be defined in terms of two values of magnetizing force, one the maximum value of magnetizing force on the particular branch with which we happen to be concerned and the other the instantaneous value of the magnetizing force on that branch. The equation for either branch may then be expressed as a double power series in these two variables, the instantaneous magnetizing force (h) which will be expressed in gilberts/cm, and the maximum value of the magnetizing force (H), expressed in the same units;

$$B(h, H) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn} h^m H^n, \tag{1}$$

where

$$a_{mn} = \frac{1}{m!\, n!} \frac{\partial B(h, H)}{\partial h^m \partial H^n} \Bigg]_{0,\,0}. \tag{2}$$

The parallelism of this representation with that for the plate current-grid potential curves of a three electrode thermionic tube is evident,—in both cases double power series are involved.[4] The coefficients a_{mn} are derivatives which are evaluated at the point $h = 0$, $H = 0$, and it will be understood in (2) that these particular values are inserted after the derivatives have been taken, in quite the usual manner.

There is a further interesting parallel here to the vacuum tube case regarding simplification of the general relation. In the case of the vacuum tube a simplification of the double power series due to van der Bijl has been employed which represents the family of tube characteristics by an equation in a single variable. This simplification consists in assuming the amplification factor to be constant, and the important point for us here is that it is equivalent to the assumption

[3] "On Hysteresis Loops and Lissajous' Figures," *Phil. Mag.*, 1910.

[4] Peterson and Evans, "Modulation in Vacuum Tubes used as Amplifiers," *Bell System Technical Journal*, July, 1927.

that the different branches of a family have the same form, so that by suitable change of plate or of grid potential the plate current-grid voltage curves may be superposed. A relation of precisely the same form in the case of magnetic hysteresis loops was stated by Lord Rayleigh [5] upon examination of data obtained by a magnetometric study of the behavior of a single low permeability specimen. Examination of his data enabled Rayleigh to conclude that the branches corresponding to different loops of a family could be superposed when referred to a common loop tip, the branches all having the same parabolic form within the limits of accuracy of the measurements, over the range of magnetizing forces involved.

It is of course evident that even if this relation held at low fields in all materials it would break down at sufficiently high fields. Further, there is no *a priori* reason to expect this relation to hold for magnetic materials other than the one Lord Rayleigh investigated unless we restrict consideration to a very small range of magnetizing force. With these ideas in mind it seems the safer procedure to assume no such simple relation between the different loops of a family, however convenient it might be, and to treat the problem in more general terms; if any such simplifying relations exist they will be made apparent after application to definite materials. We may anticipate matters a bit to state at this point that certain materials seem to obey the relation while certain others seem to violate it within the range of forces involved in communication work, and further light is shed on the significant processes involved in harmonic production by treating the problem in this way.[6]

General Equations for Hysteresis Branches. The equations of both the upper branch family and the lower branch family have the form of equation (1)—we may designate the upper branches by B_1 and the lower branches by B_2—but the coefficients of the two series differ in general.

In order to put the equations in shape so that they may be of practical utility it is now necessary to determine the coefficients of the expressions for B_1 and B_2 so that they apply to a definite loop family, and this is accomplished by reference to some of the more general properties of the loops and of the normal magnetization curve. These properties are as follows:

[5] "Notes on Electricity and Magnetism, III," *Phil. Mag.*, 1887, V. 23.

[6] Unpublished work based on assumptions which include Rayleigh's relation was independently carried out by W. P. Mason of these Laboratories in 1922 and 1923. An account of some applications of Rayleigh's relation is to be found in an interesting paper by Jordan published in the *Elektrische Nachrichten Technik*, B. 1, H. 1, July 1924.

1. When both h and H are zero the flux density on either branch is zero.
2. The flux density on one branch with H and h given is equal and opposite in sign to the flux density on the other branch corresponding to the negative of h and to the same maximum force H.
3. The two branches corresponding to a definite H meet at the normal magnetization curve.

The application of these properties to the power series enables us to deduce relations between the coefficients as demonstrated in Appendix 1:

$$\begin{aligned} a_{01} &= a_{00} = 0, \\ a_{02} &= -a_{20}, \\ a_{03} &= -a_{21}. \end{aligned} \tag{7}$$

We shall find it sufficient to include the third degree terms for our work, so that we need not investigate relations between coefficients of higher degree. The loop equations are simplified by utilizing (7) and we can make a number of interesting deductions. Thus the equation for the normal *magnetization curve* is given by

$$B(H, H) = a_{10}H + a_{11}H^2 + (a_{12} + a_{30})H^3. \tag{6a}$$

In this equation a_{10} will be recognized as the initial permeability usually expressed as μ_0 since upon division by H we have for the *permeability*

$$\mu = a_{10} + a_{11}H + (a_{12} + a_{30})H^2 \cdots.$$

According to this equation the change of permeability with magnetizing force is linear at sufficiently small fields. The above equations are, in all rigor, infinite series, but for our purposes it will be found sufficient to consider only coefficients of the third and lower orders,—in some cases the second order will suffice.

An expression for the *remanence curve* of the loop family may be obtained by setting h equal to zero in the equation for the upper branch. In that case we find from (4a)

$$B(O, H) = a_{02}H^2 + a_{03}H^3 + \cdots, \tag{8}$$

hence for sufficiently small magnetizing forces the remanence increases as the square of the magnetizing force.

The *hysteresis loss per cycle per unit volume* may also be obtained directly from the branch equations (4a) and (5a). The loss in ergs, w, is equal to the area of the hysteresis loop divided by 4π. If now we

consider the loop area to be built up of strips of infinitesimal width based on the h-axis, the height of any one of the strips is given as

$$B = B_1(h, H) - B_2(h, H)$$

and we have

$$w = \frac{1}{4\pi}\int_{-H}^{H} B dh \tag{9}$$

so that, by Appendix 1, (9) may now be expressed as

$$w = \frac{2}{3\pi}(a_{02}H^3 + a_{03}H^4 + \cdots). \tag{10}$$

It is clear, therefore, that at sufficiently low fields the hysteresis loss varies as the cube of the magnetizing force, and diverges when the field is made sufficiently large—it seems to be in general agreement with experimental results on ballistic loops and, as will be pointed out later, is verified by impedance change data under alternating excitation. In view of the remanence curve equation, (10) may be rewritten as

$$w = \frac{2}{3\pi} HB(O, H). \tag{11}$$

Various approximations have been made in the past to hysteresis loop forms in order to obtain convenient expressions for the hysteresis loss. Thus if we consider the loop as an ellipse the loss becomes

$$\frac{1}{4} HB(O, H),$$

while if we consider the loop a parallelogram the loss is

$$\frac{1}{\pi} HB(O, H).$$

Both these expressions give too large a result since the coefficient $2/3\pi$ of the exact equation (11) is 0.212.

Branch Equations for Materials Obeying Rayleigh's Relation. Rayleigh's observation enables us to establish relations between the different coefficients involved in our development above when the loops of a family are similar in form. For the sake of completeness a derivation of these relations is given in Appendix 2; their validity may then be judged by test in specific cases. In the derivation we assume a power series expansion for one branch of the largest loop of a family referred to the tip, and assume that the smaller loops are of the same form. Then by referring the equations to the origin instead of to the loop tip, we arrive at the hysteresis branch equations.

By comparing the coefficients of this equation with those of the general equation (1), we can deduce relations between the coefficients. These are, from Equation 16 of Appendix 2,

$$\begin{aligned} &a_{01} = 0, && \\ &a_{20} + a_{02} = 0, && a_{11} = 2a_{20}, \\ &a_{21} + a_{03} = 0, && a_{12} = a_{21} = 3a_{30}, \\ &a_{40} + a_{22} + a_{04} = 0, && a_{31} = a_{13} = 4a_{40} = 6a_{22}. \end{aligned} \tag{17}$$

The left column is in agreement with Equation (7), while the right column furnishes new relations between the coefficients. Thus the coefficients based on loop similarity are not inconsistent with those deduced under no assumptions, but the former involve additional relationships between coefficients which need not be satisfied in the general case.

Families of hysteresis loops obtained by the usual ballistic method have been examined for these relationships in the case of several materials with the following results for coefficients up to and including the second degree:

Hysteresis Loop Coefficients

	Silicon Steel	'B' Dust [7]	'C' Dust [7]
a_{10}	270	35.3	26.2
a_{11}	2600	1.53	0.59
$a_{11}/2$	1300	0.76	0.29_5
a_{02}	967	0.65	0.29
a_{20}	− 951	− 0.71	-0.28_5

The values tabulated were obtained by 'smoothing' data obtained from ballistic loops, which leaves a great deal to be desired from the standpoint of precision. The difference between a_{02} and $-a_{20}$ is an index of this lack of precision. The average deviation of $a_{11}/2$ from the mean of a_{02} and a_{20} (a relation deduced on the basis of similarity) is small for *C* dust and large for silicon steel while the third order coefficients show much poorer results. The lack of precision in the original data however, is such as to leave unsettled the question of loop similarity for the materials tested at the fields to which the coefficients apply; the experimental error is too great.

The use of ballistic methods for determining coefficients becomes even less satisfactory for materials with low hysteresis losses and cannot be used for analyses which have any aspirations to precision. It will be shown in the next section, however, that the significant coefficients enter into the impedance of a coil which employs the material under examination as a core, so that the desired coefficients

[7] Speed and Elmen, "Magnetic Properties of Compressed Powdered Iron," *A. I. E. E.*, V. 40.

may be obtained with the aid of AC bridge measurements under appropriate conditions for which eddy currents and winding capacities are unimportant. These measurements are of a higher order of precision than those obtained by the ballistic method, and will be treated in detail in the next section.

Part 2. Alternating Magnetization

Sinusoidal Magnetizing Force. The same magnetic characteristics which were the subject of the investigation of the last section are involved in alternating magnetization when the applied field has but one maximum and but one minimum per cycle, and when the eddy losses are not great enough to introduce screening effects. As one of the results of that investigation, we have arrived at equations for both the upper and the lower branch families, so that to obtain an expression for the flux density valid over the entire cycle, it is now necessary to combine the two branches by a Fourier's series in the usual manner.

Before doing this, however, it is necessary to express the branch equations for the flux density in terms of time, rather than as series of powers of the independent variable, the magnetizing force. To do this we substitute the equation for the magnetizing force [8]

$$h = H \cos pt \tag{18}$$

in the branch equations (4a) and (5a) of Part 1. The result of the substitution is to give the upper and lower branch flux equations in terms of powers of cos pt, which may be expressed in terms of multiple angles. These two equations are then combined in a Fourier series which is valid over the entire cycle:

$$B = \frac{b_0}{2} + \sum_{k=1}^{k=\infty} (b_k \cos kpt + a_k \sin kpt),$$

in which the coefficients are given by the usual expressions, equation 22a of Appendix 3. The coefficients b_0, b_2, $\cdots$ b_{2k}, a_0, a_2, $\cdots$ a_{2k} are found to be identically zero on account of the symmetry of the loop family about the origin, while the fundamental and third harmonic coefficients are found to be as follows:

$$\begin{aligned} a_1 &= \frac{8}{3\pi}(a_{02}H^2 + a_{03}H^3), \\ b_1 &= a_{10}H + a_{11}H^2 + (a_{12} + \tfrac{3}{4}a_{30})H^3, \\ a_3 &= -\frac{8}{15\pi}(a_{02}H^2 + a_{03}H^3), \\ b_3 &= a_{30}H^3/4. \end{aligned} \tag{25}$$

[8] The purely sinusoidal magnetizing force may be obtained, despite the varying reaction of the iron core coil, by connecting a generator through a low pass or band pass filter having a high impedance outside the pass band, or by connecting a pure sine wave generator to the coil through a high impedance.

Details of the derivation are given in Appendix 3. Inasmuch as we assume the applied field to vary as cos pt, the b's are in phase with the applied field and the a's are in quadrature. The two a's, it is observed, are connected by a constant of proportionality ($a_1 = -5a_3$) and depend upon the remanence; or upon what is the same thing, the hysteresis loss divided by the magnetizing force.

If we expanded the loop equations to higher powers we should find that a_1 and a_3 cease to be linearly proportional. The coefficient b_1 is observed to have its first three terms identical with those of the normal magnetization curve, but the fourth term differs by precisely the amount b_3.

The voltage existing across a coil enclosing a core characterized by the above coefficients is

$$E = nA10^{-8}\frac{dB}{dt}, \tag{26}$$

where n is the number of turns enclosing the core, A is the core area in cm^2, B is the total flux density and E is the total generated potential. Carrying out this operation we have

$$E = nA10^{-8}(pa_1 \cos pt + 3pa_3 \cos 3pt - pb_1 \sin pt - 3pb_3 \sin 3pt) \tag{26a}$$

in which the voltage components in phase with the current depend on the hysteresis coefficients, and the quadrature components depend only on the coefficients of the normal magnetization curve. The dependence of each of these components upon the applied field is clear from Equation (25). To take the two third harmonic components it will be observed that a_3 starts to vary with the square, while b_3 starts to vary with the cube, of the applied field. It follows therefore that at sufficiently small amplitudes the third harmonic is produced by the a_3 term and not by the b_3 term, which means that *under the conditions noted, harmonic production is due to hysteresis and not primarily to permeability change.*

From the two fundamental components of voltage across the coil an expression for the inductance and resistance offered to the flow of alternating current may be deduced. The inductance, it is easy to see, is obtained directly from the magnetization curve,—the d.c. permeability of the normal magnetization curve therefore coincides with the a.c. permeability at small fields. The resistance may be obtained from the expression derived above for hysteresis loss per cycle per cc. If we multiply that value by the volume of iron in the coil, by the frequency, and by 10^{-7} to convert ergs to watts we have the loss per second, and this may be equated to the square of the

effective fundamental current multiplied by the hysteresis resistance, or

$$wfA\pi d10^{-7} = I^2R_h/2.$$

Here d is the diameter in cm, I is the peak value of the fundamental current, f is the frequency, and R_h is the hysteresis resistance. This may be solved for the hysteresis resistance in terms of the r.m.s. value of fundamental current $\bar{I}$, which is at low fields,

$$R_h = 1.2 \cdot 10^{-8} n^3 A f \bar{I} a_{02}/d^2. \tag{27}$$

The hysteresis resistance at small fields thus varies linearly with the applied current.

It is easy to derive the relation between the hysteresis resistance and the generated third harmonic voltage at low fields since they involve the same constants a_{02} and a_{03}; from (26a) and (25) we have for the r.m.s. third harmonic voltage

$$\bar{E}_3 = 3pnA10^{-8}a_3/\sqrt{2} = 0.72 \cdot 10^{-8} f n^3 A \bar{I}^2 a_{02}/d^2.$$

Comparison with (27) shows that

$$\bar{E}_3 = 0.6 R_h \bar{I}. \tag{28}$$

This simple relation is valuable in obtaining an idea of the harmonic production in a specific coil through resistance measurements, and more than that, it enables us to determine the coefficients significant in the distorting process for the material under test, so that the harmonic production in a coil of any dimensions enclosing the same core material may be calculated. The degree of precision ordinarily attain-

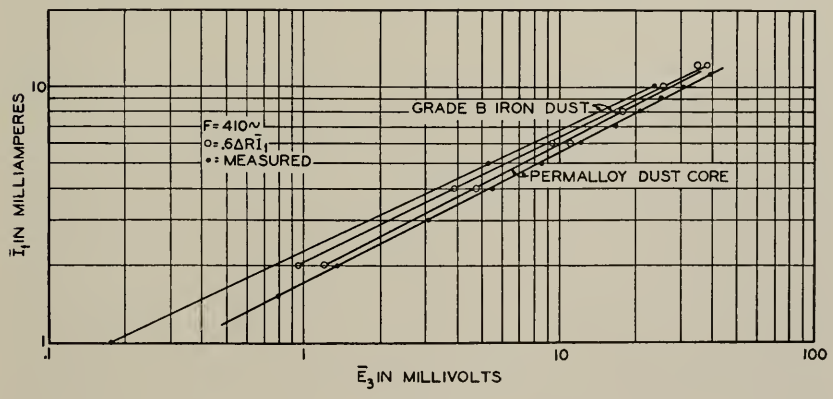


FIG. 3

able is brought out for two materials in Fig. 3, and the agreement is observed to be within the experimental error in those two representative cases.

These relations provide us with an expression for the ratio a_{11}/a_{02} which may be obtained from impedance measurements. It will be recalled that Rayleigh's relation would give this ratio the value two, and a.c. bridge measurements may be used to check this relation. The change of resistance with current is given by (27), and the change of reactance with current may be calculated from (25) and (26), since we have

$$\begin{aligned}\Delta X &= a_{11} H^2 pnA\, 10^{-8}/\bar{I}\sqrt{2} \\ &= 1.42 \cdot 10^{-8} n^3 Af\bar{I}a_{11}/d^2.\end{aligned}$$

Hence if we denote R_h by ΔR, we may write

$$\frac{a_{11}}{a_{02}} = 0.85\frac{\Delta X}{\Delta R}.$$

The results obtained on some dust cores of different composition and two solid materials may be tabulated as follows:

Material	Number of Specimens	a_{11}/a_{02}
Grade 'B' Iron Dust	10	2.65
Grade 'C' Iron Dust	1	2.63
Permalloy Dust [9] 'A'	3	2.10
" " 'B'	3	1.81
Iron Dust No. 1	1	2.80
Iron Dust No. 2	2	2.55
Perminvar [10]	1	3.0
Alloy 'D'	2	1.84

The method of analysis and the results obtained above may be summarized as follows. Starting with the general development of the hysteresis branches in a double power series and making use of only the most fundamental experimental observations, equations have been derived for the normal magnetization curve, the remanence characteristic, and the hysteresis loss characteristic in terms of the original hysteresis branch coefficients by combining the equations for the two branches in a trigonometric series. The series for the normal magnetization curve is found to start with the first power of the magnetizing force, the remanence starts with the second power, and the hysteresis loss starts with the third power, the curves varying in accordance with their respective first terms when H is small. These results seem to be in general agreement with experience. For large values of H the shapes of the different curves depend of course upon the size and sign of the coefficients involved, about which nothing can be said until the coefficients have been evaluated from experimental data.

[9] Shackelton and Barber, "Compressed Powdered Permalloy," A. I. E. E. Convention, February 1928.

[10] Elmen, "Magnetic Properties of Perminvar," *J. F. I.*, September 1928.

In the case of a sinusoidal magnetizing force the fundamental and third harmonic flux densities were derived with the aid of the trigonometric series development. These were shown to depend in a simple way at low fields upon the normal magnetization and hysteresis loss characteristics. The in-phase fundamental voltage component depends directly upon the hysteresis loss per unit current, while the fundamental component in quadrature varies with the magnetization curve at low forces. The in-phase third harmonic varies with the in-phase fundamental at low forces, and the third harmonic in quadrature comes in only with larger forces in a manner which depends upon the magnetization characteristic. The range of forces over which our equations are valid depends simply upon the number of terms taken in the development of the branch equations, and is not necessarily restricted to small forces. The hysteresis loop coefficient enters into the impedance offered to the fundamental frequency by a coil enclosing the core material in question in such a way that the third harmonic produced may be deduced from the change of resistance with current. Further, the ratio of the change of reactance with current to the change of resistance with current may be used to provide a test of Rayleigh's relation, which is found to hold for some materials, while it is invalid for others. The precision obtainable in the evaluation of hysteresis loop coefficients is much greater by the a.c. bridge measurement under proper experimental conditions than by the analysis of ballistic loops. Incidentally attempts have been made to obtain *B–H* loops by *AC* methods with the aid of the Braun tube, for example,[11] but the precision attainable is not sufficiently high for our purpose.

Complex Magnetizing Force. Our preceding analysis has furnished us with the fundamental voltage drop across a coil enclosing the iron core under consideration, together with the third harmonic voltage generated in the coil winding due in general partly to the non-linear *B–H* relation and partly to the effect of hysteresis. The generated voltages corresponding to the fifth, seventh, and higher orders are also calculable by the same methods but will not be specifically considered since they are smaller than the third harmonic at low fields. This generated third harmonic voltage exists in its entirety across the coil winding only when the impedance of the external circuit is much higher than that of the coil at the harmonic frequency, a condition not usually satisfied in telephone circuits. A current of the third harmonic frequency then flows in the circuit, its amplitude and phase depending evidently upon the generated third harmonic voltage—that is, the coil structure and core material—as well as upon the total

[11] Peterson, *Phys. Rev.*, Vol. 27, No. 3, p. 320.

circuit impedance to the third harmonic which includes that of the coil. We have now to determine the coil impedance to a complex magnetizing force constituted by the fundamental and the third harmonic. It is evident at the outset that if we restrict consideration to a very small third harmonic, the impedance to the fundamental cannot be very materially altered.

In general the fundamental and third harmonic may exist in any phase but for our present purpose we shall take the magnetizing force to be

$$H = H_1 \cos pt + H_3 \cos 3pt, \tag{29}$$

in which the phase angle is assumed zero. This seems to be an arbitrary assumption, but it may be shown that the results are not materially different for other phase displacements, and there is some gain toward simplicity of treatment by taking the phase angle zero. The equations of Part 1 may then be carried over without change, and details of the analysis are given in Appendix 4.

The ratio of the resistances to the third harmonic and to the fundamental is 0.77, from the relations deduced in the appendix. Some experimental work has been done to test the validity of the expressions derived above for the reactance and resistance components of the third harmonic, the results of which are given in Fig. 4. It is seen that the check for the inductance to the third harmonic is very good but that the resistance component appears to be substantially that of the fundamental rather than 77 per cent of it, as the analysis indicates.

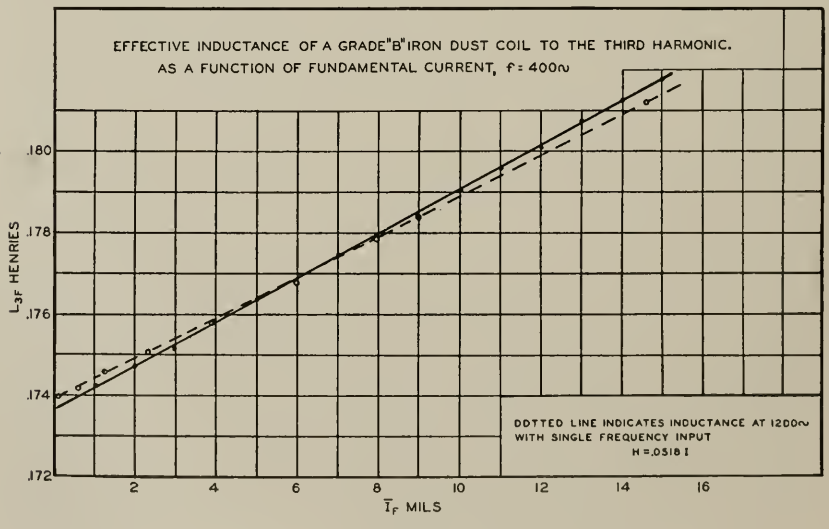


FIG. 4A

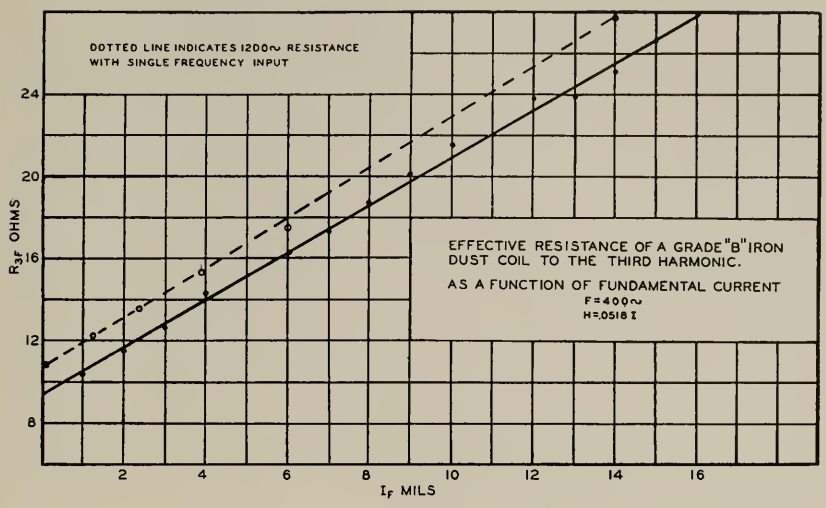


FIG. 4B

The experimental method for obtaining this result—the impedance to a small third harmonic in the presence of a relatively large fundamental—is illustrated in Fig. 5. The method consists in measuring the third harmonic current by means of a current analyzer [12] for a number of circuit conditions in which the fundamental current is

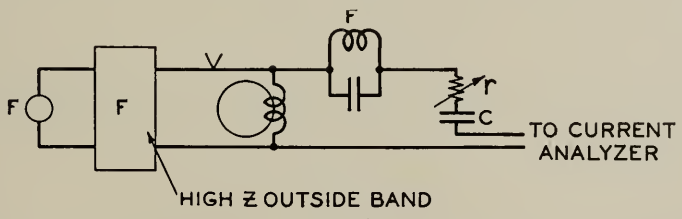


FIG. 5

maintained constant. The circuit is first tuned to the third harmonic by varying the capacity C in the third harmonic circuit, and the current is then measured for a series of values of the series resistance r. A shunt resonant circuit tuned to the fundamental is inserted in the third harmonic path so as to separate effectively the third harmonic circuit from that of the fundamental. With this precaution taken to avoid harmonic production in the analyzer and to maintain the fundamental current constant while r and C are varied, the inductance to the third harmonic is obtained from the resonating capacity, and the resistance is determined as that value of r for which the third harmonic current falls to half its maximum.

[12] For details of current analysis see the paper by A. G. Landeen, *B. S. T. J.*, April 1927.

Part 3. Application of the Analysis

Air-Gaps and Dilution. Under certain conditions improvement in the operation of iron core coils toward freedom from harmonic production may be attained by inserting air-gaps in the magnetic path. The expressions which we have derived up to this point are valid for a material having the constants assigned, and the question now arises as to the parameters which characterize the operation of the iron core including air-gaps, and their relation to the parameters for the original core without air-gaps. With these relations given, our previous work may be applied to cores with air-gaps.

In establishing the correspondence between the parameters for the two cases, it is instructive to use two methods—one a direct attack,[13] the other resting on an analogy with non-linear vacuum tube circuits.[14] We may determine the effects sought for by the direct method on consideration of a single branch of a hysteresis loop, which is expressed by equations (4*a*) or (5*a*) of Part 1,

$$B(h, H) = \sum\sum a_{rs} h^r H^s. \tag{36}$$

Now with an air-gap in the magnetic circuit, the magnetomotive force effective is that applied, reduced by the drop across the air-gap, or

$$\begin{aligned} m' &= m - \rho\varphi, \\ M' &= M - \rho\Phi, \end{aligned} \tag{37}$$

in which mM, $\varphi\Phi$ are instantaneous and maximum values, respectively, of the impressed m.m.f. and flux, and in which ρ represents the air-gap reluctance

$$\rho = \lambda/A, \tag{38}$$

λ being the length of air-gap and A the core cross-section.

In order to apply (37) we re-express (36) in terms of magnetomotive force and flux as follows:

$$\varphi(m, M) = \sum_r \sum_s \frac{A a_{rs}}{l^{r+s}} m'^r M'^s, \tag{39}$$

where l is the length of magnetic circuit in the iron. Equations (37) may now be substituted in (39) to yield

$$\varphi(m, M) = \sum_r \sum_s \frac{A a_{rs}}{l^{r+s}} (m - \rho\varphi)^r (M - \rho\Phi)^s. \tag{40}$$

[13] Due to Mr. H. P. Evans.

[14] See Appendix 5.

This equation may be solved for φ by identifying coefficients when we substitute the solution

$$\varphi(m, M) = \sum_r \sum_s \frac{A a_{rs}'}{l^{r+s}} m^r M^s,$$

which is equivalent to

$$B(h, H) = \sum_r \sum_s a_{rs}' h^r H^s. \tag{41}$$

Carrying out the operations, the primed coefficients are determined in terms of the original coefficients and the core structure as follows:

$$\begin{aligned} a_{10}' &= a_{10} \Big/ \left(1 + \frac{\lambda}{l} a_{10}\right), \\ a_{02}' &= -a_{20}' = a_{02} \Big/ \left(1 + \frac{\lambda}{l} a_{10}\right)^3, \\ a_{11}' &= a_{11} \Big/ \left(1 + \frac{\lambda}{l} a_{10}\right)^3. \end{aligned} \tag{42}$$

It is clear that the effect of an air-gap is to diminish the higher order coefficients to a greater extent than those of lower order; no new coefficients are introduced. If we refer to a constant flux density, then the impressed force is

$$H' = H\left(1 + \frac{\lambda}{l} a_{10}\right) \tag{43}$$

and the modulation voltage becomes proportional to

$$a_{02}' H'^2 = \frac{a_{02} H^2}{1 + \frac{\lambda}{l} a_{10}},$$

so that the harmonic e.m.f. has been reduced in the ratio of $1 + \frac{\lambda}{l} a_{10}$ to unity, which is precisely the ratio of initial permeabilities.

Dilution. If the magnetic material is diluted with a non-magnetic substance (insulation) so that the effective air-gap length is λ, the above equations for the flux density still hold. The effective cross-sectional area, however, is reduced to

$$A' = A\left(\frac{l - \lambda}{l}\right)^2, \tag{44}$$

which is ordinarily of small account compared to the other factors involved.

Choice of Core Material. The ideal characteristics for a core ma-

terial in respect to its freedom from harmonic production may be determined by a consideration of the circuit problem. The third harmonic driving e.m.f. is obtained from the last section as

$$E_3 = pnAa_{02}H^2 10^{-8}, \\ \doteq \frac{n^3 A}{d^2} a_{02} I_1^2, \tag{45}$$

in which n represents the number of turns enclosing the core of area A, d is the mean diameter of the toroidal core and a_{02} is the hysteresis coefficient.

Suppose that we have to design a coil of fixed inductance in which the harmonic is to be a minimum, so that

$$L = K\frac{n^2 \mu A}{d} = \text{const.} \tag{46}$$

We proceed to the consideration of a number of special cases subject to condition (46).

Case 1—L **Fixed,** *n* **Variable.** From (46)

$$n = \text{const.} \left(\frac{d}{\mu A}\right)^{1/2},$$

which may be substituted in (45) to give

$$E_3 \doteq \frac{1}{\sqrt{Ad}} \frac{a_{02}}{(\mu)^{3/2}}. \tag{47}$$

Hence in a coil of fixed inductance, fixed core area, and fixed core diameter, minimum harmonic voltage is produced with a material for which $a_{02}/\mu^{3/2}$ is minimum.

Case 2—L **Fixed,** *A* **Variable.** From (46)

$$A = \text{const. } d/n^2 \mu,$$

which, upon insertion in (45), yields

$$E_3 \doteq \frac{n}{d} \frac{a_{02}}{\mu}, \tag{48}$$

in which the modulated voltage varies linearly with a_{02}/μ.

Case 3—L **Fixed,** *d* **Variable.** By a similar procedure, we obtain an expression for the generated third harmonic voltage

$$E_3 \doteq \frac{1}{nA} \frac{a_{02}}{\mu^2}. \tag{49}$$

For further work we suppose the coil reactance considerably greater

than the circuit resistance at the fundamental frequency, and consider the case of an input transformer in which the secondary is practically open-circuited.

Case 4—**Input Transformer.** In accordance with the above assumption the fundamental current through the coil is determined by the coil reactance x or

$$I_1 = \frac{E}{x} \doteq \frac{d}{n^2 \mu A}.$$

Putting this in (45), we find

$$E_3 \doteq \frac{1}{nA} \frac{a_{02}}{\mu^2}, \tag{50}$$

which is identical with Equation (49).

In the four cases treated above it has appeared that there are three quantities which characterize the modulating properties of a ferromagnetic coil under different conditions—a_{02}/μ, $a_{02}/\mu^{3/2}$, a_{02}/μ^2. The values of these ratios have been determined for a number of materials and are tabulated below.

	a_{02}/μ	$a_{02}/\mu^{3/2}$	a_{02}/μ^2
B dust	18×10^{-3}	3.1×10^{-3}	0.52×10^{-3}
C dust	11	2.2	0.42
Permalloy dust [15]	17	2.1	0.24
Perminvar [15]	2.1	0.1	0.0045
Silicon Steel	2500	210	16

These results apply only when eddy currents are negligible; as the frequency is raised eddy currents effectively screen the core and the harmonic flux is reduced to a greater extent than is the fundamental. This effect, as is well known, depends upon the specific resistivity of the core material and will not be evaluated here.

It has been pointed out that under different circuit conditions an index of the modulation is given by the ratios a_{02}/μ, $a_{02}/\mu^{3/2}$, a_{02}/μ^2, depending upon specific conditions. For diluted materials, these ratios referred to the undiluted material become

$$\frac{a_{02}}{a_{10}\left(1 + \frac{\lambda}{l} a_{10}\right)^2}, \quad \frac{a_{02}}{\left[a_{10}\left(1 + \frac{\lambda}{l} a_{10}\right)\right]^{3/2}}, \quad \frac{a_{02}}{a_{10}^2\left(1 + \frac{\lambda}{l} a_{10}\right)},$$

and the utility of air-gaps toward the reduction of harmonic is made evident quantitatively in every case.

In order to test these relations a comparison was made of the

[15] Laboratory specimens.

coefficients for specimens of grade "B" and grade "C" dust which represent different dilutions of electrolytic iron. The coefficients used were those tabulated in Part 1. In accordance with equations (42) above we may write

$$\begin{aligned} a_{10}\Big/\left(1+\frac{\lambda_1}{l}a_{10}\right) &= 26.2, & a_{10}\Big/\left(1+\frac{\lambda_2}{l}a_{10}\right) &= 35.3, \\ a_{11}\Big/\left(1+\frac{\lambda_1}{l}a_{10}\right)^3 &= 0.59, & a_{11}\Big/\left(1+\frac{\lambda_2}{l}a_{10}\right)^3 &= 1.53, \\ a_{02}\Big/\left(1+\frac{\lambda_1}{l}a_{10}\right)^3 &= 0.29, & a_{02}\Big/\left(1+\frac{\lambda_2}{l}a_{10}\right)^3 &= 0.65. \end{aligned} \tag{51}$$

From these, we should have equality of the ratios

$$0.59/1.53, \qquad 0.29/0.65, \qquad (26.2/35.3)^3,$$

or

$$0.37, \qquad 0.45, \qquad 0.41,$$

which are evidently in fair agreement. It is clear then that the properties of diluted materials may be calculated from the characteristics of the original material, at least when the process of dilution does not change the intrinsic properties of the magnetic material involved.

Applications to Transformers. The third harmonic flux component may be obtained when the fundamental magnetizing force is given, and the latter is easily obtained in toroidal core inductance coils from the relation

$$h = 0.4ni/d, \tag{52}$$

where both h and i refer to the fundamental frequency. In a transformer, however, the net magnetizing force is obtained as the sum of two components, one due to the primary and the other produced by the secondary, so that some further investigation is required before the net magnetizing force in the core may be calculated in terms of the transformer constants and the primary current.

Net Magnetizing Force. To obtain the net magnetizing force which determines uniquely the generated flux components we are required to obtain the primary and secondary currents ($i_1 i_2$), to multiply each current by the number of times it encircles the core, to add the two products, and finally to multiply the sum by $0.4/d$:

$$H_{\text{net}} = 0.4(n_1 i_1 + n_2 i_2)/d. \tag{53}$$

To evaluate (53) then we shall solve for the primary and secondary fundamental currents in terms of the circuit constants and the applied potential.

The transformer circuit equations from which the currents may be evaluated are the familiar ones

$$\begin{aligned} E &= Z_1 i_1 + jpMi_2, \\ O &= Z_2 i_2 + jpMi_1, \end{aligned} \tag{54}$$

in which we assume the transformer to be an idealized structure—the capacity and leakage effects of actual transformers will be assumed combined with the connected impedances for simplicity—they will therefore not be dealt with explicitly. From the second of (54) is obtained the relation between the two currents

$$i_2 = -jpMi_1/Z_2 \tag{55}$$

which may be substituted in the first of (54) to give the usual expression for the primary current

$$i_1 = E \Big/ \left[R_1 + \frac{p^2M^2}{Z_2^2} R_2 + j \left(X_1 - \frac{p^2M^2}{Z_2^2} X_2 \right) \right]. \tag{56}$$

An expression for the net magnetizing force in terms of i_1 may be had putting (55) and (56) in (53)

$$H_{\text{net}} = \frac{0.4n_1 i_1}{d} \left(1 - \frac{jpMK}{Z_2} \right), \tag{57}$$

in which K represents the turns ratio:

$$K = \frac{n_2}{n_1} = \frac{\sqrt{X_2}}{\sqrt{X_1}}. \tag{58}$$

If we make the convenient assumption, which is closely approximated under working conditions, that the coupling is perfect (no leakage) we have

$$pM = \sqrt{X_1 X_2} \tag{59}$$

and (57) may be simplified to the form

$$H_{\text{net}} = \frac{0.4n_1 i_1}{d} \frac{R_2}{Z_2}. \tag{60}$$

This equation shows that the net magnetizing force may be obtained from that of the primary alone by applying the reduction factor R_2/Z_2. Now in well designed transformers the winding reactances are much greater than the resistances to which they are connected, and in this case

$$X^2_{1,\,2} \gg R^2_{1,\,2}. \tag{61}$$

Applying this further simplification to (60) we have finally for the net

magnetizing force

$$H_{\text{net}} = \frac{0.4n_1i_1}{d}\frac{R_2}{X_2}, \tag{60a}$$

in which the reduction factor applied to the force due to the primary alone is seen to be the ratio of resistance to reactance in the secondary circuit.

This equation with the aid of (56) may now be put in terms of the primary voltage—a fixed quantity—rather than in terms of the primary current which is variable according to the circuit resistances used. Applying the assumptions (59) and (61), Equation (56) may be written

$$i_1 = E/(R_1 + R_2/K^2), \tag{56a}$$

which may be put in (60*a*) to express the net force explicitly in terms of the circuit constants:

$$H_{\text{net}} = \frac{0.4n_1E}{dX_1}\left(\frac{1}{1 + K^2R_1/R_2}\right) \tag{62}$$

In view of (58). When the transformer is terminated in its normal resistances the expression within parentheses reduces to the value one-half. It is of interest in this connection to compare the field here with that produced with the secondary open; in the latter case it is clearly

$$0.4n_1E/dX_1,$$

which is twice as great as the field in the properly terminated transformer. This result is otherwise evident, for in the properly terminated transformer but half the generator voltage is applied across the primary.

Third Harmonic. According to Equation (28) the amplitude of third harmonic voltage generated in a coil of n turns encircling a core subjected to a magnetizing force of H_{net} gilberts/cm is

$$E_3 = \frac{8}{5\pi}10^{-8}a_{02}pnAH^2_{\text{net}} \tag{63}$$

and the generated primary or secondary voltage is obtained by attaching appropriate subscripts to n. Because of the coupling, the two circuits react on one another at the third harmonic frequency as they did at the fundamental frequency, and to find the two third harmonic currents we must go through a procedure analogous to that followed for the fundamentals except that here we have an independent generator voltage effective in each of the two windings.

Indicating quantities referring to the third harmonic frequency

by the subscripts p, s for primary and secondary circuits, respectively, we have the circuit equations for the third harmonic currents:

$$\begin{aligned} e_p &= Z_p i_p + j3pMi_s, \\ Ke_p &= Z_s i_s + j3pMi_p, \end{aligned} \tag{64}$$

and by equating the two expressions for e_p we obtain a relation between the two currents:

$$i_s = \frac{KZ_p - j3pM}{Z_s - j3pMK} i_p. \tag{65}$$

Putting i_p in terms of e_p, and using (59) and (61) we find

$$i_s = e_p K R_p / R_s X_p (1 + K^2 R_p / R_s). \tag{66}$$

Suppose now that we are dealing with pure resistance loads so that

$$R_1 = R_p, \qquad R_2 = R_s.$$

We may then substitute (63) for e_p in (66) to obtain the third harmonic current in the secondary.

From (63)

$$e_p = \frac{8}{5\pi} 10^{-8} a_{02} p n_1 A \frac{0.16 n_1^2 E^2}{d^2 X_1^2} \left(\frac{1}{1 + K^2 R_1 / R_2} \right)^2, \tag{67}$$

and by substitution in (66)

$$i_s = \frac{E^2 K R_1 \alpha}{3 X_1 R_2 \left(1 + K^2 \frac{R_1}{R_2} \right)^3}. \tag{68}$$

If we consider the factor containing the resistances,

$$F = \frac{R_1}{R_2} \frac{1}{(1 + K^2 R_1 / R_2)^3}, \tag{69}$$

it is zero for R_1/R_2 zero or infinite, and reaches a maximum under the condition

$$\frac{\partial F}{\partial (R_1/R_2)} = 0. \tag{70}$$

The third harmonic secondary current is maximum when

$$K^2 = \frac{R_2}{2R_1}. \tag{71}$$

Now K^2 is fixed by the turns ratio (58), so that we may say the secondary third harmonic current is maximum when the primary resistance is made half its nominal value, or when the secondary

resistance is made twice its nominal value. This is well borne out by Fig. 6 due to A. G. Landeen taken under conditions to which the above discussion applies.

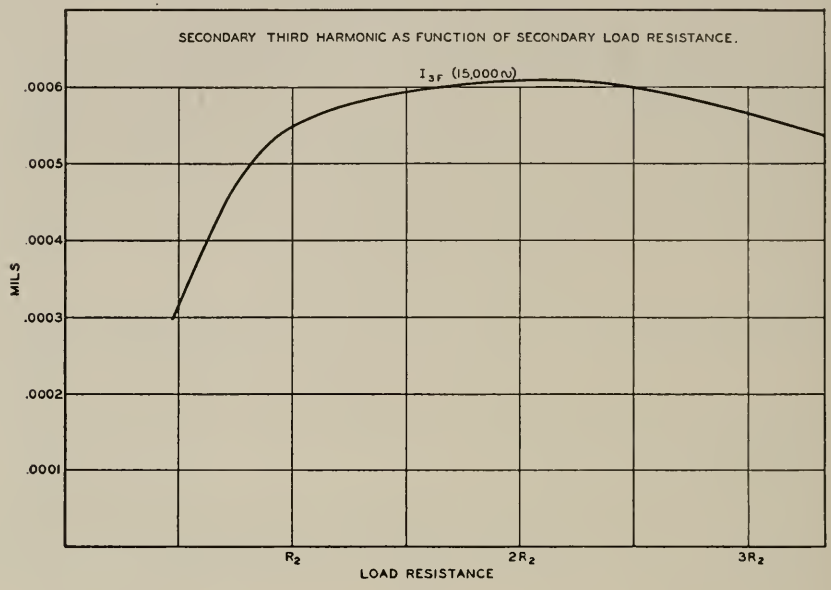


FIG. 6

At first sight the result obtained, in which the secondary harmonic current increases as the secondary resistance is increased for all values of secondary resistance below twice the nominal secondary resistance, seems a bit strange since the fundamental secondary current decreased under the same conditions. A little thought however shows that the increase of secondary resistance acts in two ways; first to reduce the harmonic current directly, and second to increase the net magnetizing force in the core which in turn increases the generated harmonic. Thus with zero secondary resistance the net magnetizing force is zero and no harmonic flux can be produced, while with large secondary loads the flux density in the core or the net magnetizing force reaches an asymptotic maximum so that the generated third harmonic increases slowly if at all while the reduction in secondary current by the resistance is continuously increasing. Under the assumptions noted Equation (68) is valid for any circuit condition and it will be observed that i_s decreases with the primary resistance below the optimum point. The value of α is explicitly

$$\alpha = 2.72 \times 10^{-10} \frac{a_{02}}{p^2 L_1{}^3} \left(\frac{n_1{}^3 A}{d^2} \right), \tag{72}$$

in which the bracketed expression coincides with the variable factor in equation (45). Hence when L_1 is constant the choice of core material follows the rules laid down for inductance coils. When the inductance is not constant, however, the factor becomes proportional to

$$\alpha = \text{const.} \frac{da_{02}}{n_1^3 \mu^3 A^2} \tag{73}$$

with the understanding that the turns ratio, resistances and frequency are fixed. Thus the secondary harmonic current with a given material is reduced by increasing the turns and core area and reducing the diameter. As far as core materials go, we have for the significant ratio a_{02}/μ^3 the values:

Material	a_{02}/μ^3
Silicon Steel	130×10^{-4}
Permalloy Dust [16]	2.4×10^{-4}
B Dust	5.3×10^{-4}
C Dust	4.3×10^{-4}
Perminvar [16]	0.05×10^{-4}

The great superiority of perminvar indicated above is restricted of course to fields of the order of less than 0.1 gilbert/cm. above which it tends to become smaller; at a field of 0.7 for example the above factor for perminvar would be multiplied by the factor three. Permalloy dust is observed to be approximately twice as good as the iron dust cores.

A very important question to answer with regard to transformer cores is this—what benefit can be gained as to harmonic production by inserting air-gaps, or by diluting the core material, leaving everything else unchanged. This is evidently to be answered by an investigation of the ratios a_{02}/μ^3 and a'_{02}/μ'^3, the primes referring as before to the diluted material. These relations are given by Equation (42) in which μ and a_{10} are interchangeable:

$$a_{10}' = a_{10} \Big/ \left(1 + \frac{\lambda}{l} a_{10}\right) = \mu',$$

$$a_{02}' = a_{02} \Big/ \left(1 + \frac{\lambda}{l} a_{10}\right)^3.$$

These permit us to evaluate a'_{02}/μ'^3:

$$\frac{a_{02}'}{\mu'^3} = \frac{a_{02}}{\left(1 + \frac{\lambda}{l} a_{10}\right)^3} \frac{\left(1 + \frac{\lambda}{l} a_{10}\right)^3}{a_{10}^3} = \frac{a_{02}}{\mu^3}, \tag{74}$$

[16] Laboratory specimens.

which clearly indicates that no change is produced in the secondary third harmonic current by introducing air-gaps or by diluting the core material, the change in core area being negligible. This relation has been verified experimentally.

The discussion above considers a constant potential applied to a transformer circuit and the relations derived are somewhat changed when we consider constant current or constant potential transformers. We seldom have to do with constant current transformers but constant potential transformers are met with frequently in vacuum tube circuits as input transformers or interstage transformers. The above discussion may be applied to this case by taking the primary generator resistance much lower than its nominal value—a condition, incidentally, which works toward the suppression of harmonic.

The conclusions are also somewhat altered when we have networks offering different impedances to the harmonic and the fundamental. Thus it is evident that the third harmonic current can be eliminated from both primary and secondary circuits by inserting a high series impedance to the harmonic frequency, or by encircling the core with a winding connected to a network which has a very low impedance to the third harmonic and high impedance to the fundamental. Further, in single frequency transmission the result may be equally well obtained by shunting a series tuned circuit around the primary or secondary winding.

Applications to Harmonic Production in Loaded Lines

If distortion takes place at any one point of a loaded line, the amount of distorted current received at the far end depends upon the amplitudes of the currents producing it, and upon the line attenuation from the point of origin to the far end. The phase similarly depends upon the phase of the fundamentals, and upon the phase shift of the line. When we have a number of distorting sources at different points along the line, the net distorted output is obtained by combining vectorially the currents due to the individual centers of distortion, since it may be assumed that no interaction exists. In a uniformly loaded line we may think of these sources of distortion as being uniformly distributed, but the amount of distortion introduced is, on the contrary, not distributed uniformly on account of the line attenuation which reduces the distortion generated at the far end of the line.

It is apparent, then, that if we are to calculate the net distortion introduced by the line, a complete specification of the phase shift and attenuation is required, together with a knowledge of the law of production of the distortion. This last also requires a specification of

amplitude and phase angle in their relation to the fundamentals producing the distortion.

A situation somewhat analogous to the one under discussion which is less involved, however, is found in the diffraction of light, in which the illumination at a point proceeds from a number of coherent sources at various distances from the point. An important difference in the two cases is the attenuation of the transmitting medium which is ordinarily small in the optical case, so that the results of the optical investigations cannot be taken over directly.

In the following we propose to calculate the third harmonic output currents with the aid of our previously established relations. The law of harmonic production has been quite definitely established in the low frequency-low flux density range. The driving third harmonic voltage of frequency $3f$ produced by a fundamental current of *rms* value $\bar{I}$ is given by (25) and (26*a*) as

$$E_3 = 0.72 \times 10^{-8} a_{02} \frac{n^3 A}{d^2} f \bar{I}^2 = M \bar{I}^2, \tag{75}$$

while the phase angle of the third harmonic generated potential is related to that of the fundamental current by the expression

$$\theta_3 = 3\theta_1. \tag{76}$$

The above data are sufficient for a solution of the problem of distortion in continuously loaded lines when the propagation constant is known as a function of frequency.

The fundamental current at any point distant x from the sending end of the continuously loaded, properly terminated line is

$$I = I_0 \epsilon^{-Px}, \tag{77}$$

where

$$P = \alpha + j\beta.$$

The third harmonic driving e.m.f. dE_3 generated in a length dx of the line may be written with the aid of (75), (76), and (77) as

$$dE_3 = M I_0^2 \epsilon^{-(2\alpha + j3\beta)x}. \tag{78}$$

Writing the line impedance as Z_0, the resulting current at x is

$$dI_3 = \frac{M I_0^2}{2Z_0} \epsilon^{-(2\alpha + j3\beta)x}. \tag{79}$$

With the line parameters a function of frequency we may write for the propagation of third harmonic current

$$i_3 = I_3 \epsilon^{-(\gamma + j\delta)x}. \tag{80}$$

The distortion at x produces a third harmonic current at the receiving end

$$di_3 = \frac{MI_0^2}{2Z_0} \epsilon^{-(2\alpha+j3\beta)x} \epsilon^{-(\gamma+j\delta)(l-x)}. \tag{81}$$

The total third harmonic current at the receiving end may now be obtained by integrating (81) over the length of the line l, which gives us

$$i_3 = \frac{MI_0^2}{2Z_0} \frac{\epsilon^{-(\gamma+j\delta)l}}{[(2\alpha - \gamma) + j(3\beta - \delta)]} (1 - \epsilon^{-[2\alpha-\gamma+j(3\beta-\delta)]l}). \tag{82}$$

Inasmuch as we are concerned with the output amplitude, the above expression may be put in a somewhat more convenient form:

$$|i_3|^2 = \left(\frac{MI_0^2}{2Z_0}\right)^2 \frac{\epsilon^{-2\gamma l}}{(2\alpha - \gamma)^2 + (3\beta - \delta)^2} \times [1 + \epsilon^{-2(2\alpha-\gamma)l} - 2\epsilon^{-(2\alpha-\gamma)l} \cos (3\beta - \delta)l]. \tag{83}$$

Thus when l is zero the harmonic vanishes as it should, and as l increases the current passes through maxima and minima determined by the cosine term. If the attenuation is not very great, the maxima occur approximately at the line lengths

$$l = \frac{(2n - 1)\pi}{3\beta - \delta},$$

where n is a positive integer. These distances correspond to odd half wave-lengths, as is true of the optical case. As l is increased the bracketed expression approaches unity and the current falls exponentially. Before this point is reached, however, the current increases in certain regions as the line length is increased. The fact that in an actual line the parameters vary with frequency means that these maxima and minima will vary in position according to the frequency, so that a maximum for one frequency may well coincide with a minimum for another frequency.

In the case of lumped loading, the integrations used for continuous loading are replaced by summations, as was done by Mason in the unpublished investigation previously cited.[4] If the spacing between coils is x_1, and if we have n coils for which $nx_1 = y$, the received third harmonic current at the end of a properly terminated line may be written

$$i_3 = \frac{MI_0^2}{2Z_0} \frac{\epsilon^{-(\gamma+j\delta)l}\{1 - \epsilon^{-[2\alpha-\gamma+j(3\beta-\delta)]y}\}}{(\epsilon^{[2\alpha-\gamma+j(3\beta-\delta)]x_1} - 1)}, \tag{84}$$

which is to be compared with (82) for the continuously loaded line.

Appendix 1: Simplification of Loop Equations

From the first of the three properties mentioned we have $a_{00} = 0$, and from the second property

$$B_1(h, H) = -B_2(-h, H) \tag{3}$$

whence, if we write

$$\begin{aligned} B_1(h, H) = {} & a_{10}h + a_{01}H \\ & + a_{20}h^2 + a_{11}hH + a_{02}H^2 \\ & + a_{30}h^3 + a_{21}h^2H + a_{12}hH^2 + a_{03}H_3 \cdots, \end{aligned} \tag{4}$$

then

$$\begin{aligned} B_2(h, H) = {} & a_{10}h - a_{01}H \\ & - a_{20}h^2 + a_{11}hH - a_{02}H^2 \\ & + a_{30}h^3 - a_{21}h^2H + a_{12}hH^2 - a_{03}H^3 \cdots. \end{aligned} \tag{5}$$

From the third property, the two branches meet at the loop tip which lies on the magnetization curve for which $h = H$, or

$$B_1(H, H) = B_2(H, H). \tag{6}$$

From the two equations (4) and (5) we have by virtue of this relation

$$a_{01}H + (a_{20} + a_{02})H^2 + (a_{21} + a_{03})H^3 = 0$$

and since this relation holds for every value of H, the coefficients of each power of H must be zero so that

$$\begin{aligned} a_{01} &= 0 = a_{00}, \\ a_{02} &= -a_{20}, \\ a_{03} &= -a_{21}. \end{aligned} \tag{7}$$

The final expressions for the branches are then evidently obtained by putting (7) in (4) and (5) which become

$$\begin{aligned} B_1(h, H) = {} & a_{10}h \\ & - a_{02}h^2 + a_{11}hH + a_{02}H^2 \\ & + a_{30}h^3 - a_{03}h^2H + a_{12}hH^2 + a_{03}H^3, \end{aligned} \tag{4a}$$

$$\begin{aligned} B_2(h, H) = {} & a_{10}h \\ & + a_{02}h^2 + a_{11}hH - a_{02}H^2 \\ & + a_{30}h^3 + a_{03}h^2H + a_{12}hH^2 - a_{03}H^3. \end{aligned} \tag{5a}$$

Appendix 2: Rayleigh's Relation

If we suppose the lower branch of any loop, when referred to the tip of the largest loop considered, to be given by the equation

$$B_1 = \mu_0 h_1 + \nu h_1^2 + \lambda h_1^3 + \omega h_1^4 \tag{12}$$

we may refer the family of branches to the origin by the transformation

$$\begin{aligned} h &= h_1 - H, \\ B &= B_1 - B_m, \end{aligned} \tag{13}$$

if $B_m H$ refer the midpoint of the largest loop to the tip. Then

$$B_m = \mu_0 H + 2\nu H^2 + 4\lambda H^3 + 8\omega H^4. \tag{14}$$

Putting (13) in (12)

$$B + B_m = \mu_0(h + H) + \nu(h + H)^2 + \lambda(h + H)^3 + \omega(h + H)^4,$$

whence, subtracting (14),

$$\begin{aligned} B' = \mu_0 h + \nu(h^2 + 2hH - H^2) + \lambda(h^3 + 3h^2H + 3hH^2 - 3H^3) \\ + \omega(h^4 + 4h^3H + 6h^2H^2 + 4hH^3 - 7H^4), \end{aligned} \tag{15}$$

which represents the hysteresis branch equation referred to the origin, on the basis of loop similarity.

The coefficients obtained by the two methods may now be compared. Thus identifying coefficients of (15) with those of the general equation (1)

$$\begin{array}{llll} a_{10} = \mu_0, & a_{11} = 2\nu, & a_{02} = -\nu & a_{12} = 3\lambda, \\ a_{20} = \nu, & a_{21} = 3\lambda, & a_{03} = 3\lambda, & a_{13} = 4\omega, \\ a_{30} = \lambda, & a_{31} = 4\omega, & a_{04} = -7\omega, & a_{22} = 6\omega. \\ a_{40} = \omega, & & & \end{array} \tag{16}$$

Appendix 3: Alternating Magnetization, Sinusoidal Magnetizing Force

The resulting expression is simplified if we make the following substitutions

$$\begin{aligned} \alpha &= a_{02}H^2 + a_{03}H^3 = B(O, H), \\ \beta &= a_{10} + a_{11}H + a_{12}H^2, \\ \delta &= a_{30}H^3. \end{aligned} \tag{19}$$

It may be noted that α is the remanence and that β is an approximation to the permeability, in fact the permeability is given as the sum of β and δ. With (18) and (19) inserted in the branch equations, then, we have

$$\begin{aligned} B_1(H\cos pt, H) &= \alpha + \beta\cos pt - \alpha\cos^2 pt + \delta\cos^3 pt, \\ B_2(H\cos pt, H) &= -\alpha + \beta\cos pt + \alpha\cos^2 pt + \delta\cos^3 pt. \end{aligned}$$

For convenience we shall express these relations in terms of multiple angles, and we have for the equation of the upper loop family

$$B_1(H\cos pt, H) = -\alpha/2 + (\beta + 3\delta/4)\cos pt - \alpha/2\cos 2pt + \delta/4\cos 3pt.$$

If we write

$$\begin{aligned} A &= \alpha/2 = (a_{02}H^2 + a_{03}H^3)/2, \\ B &= \beta + 3\delta/4 = a_{10} + a_{11}H + a_{12}H^2 + 3a_{30}H^3/4, \\ C &= \quad = -A, \\ D &= \delta/4 = a_{30}H^3/4, \end{aligned} \tag{20}$$

the final form for the loop equations is

$$\begin{aligned} B_1(H\cos pt, H) &= \quad A + \mathrm{B}^{17}\cos pt + C\cos 2pt + D\cos 3pt, \\ B_2(H\cos pt, H) &= -A + B\ \cos pt - C\cos 2pt + D\cos 3pt. \end{aligned} \tag{21}$$

We are now in position to combine the two equations of (21) in a Fourier series valid over the entire cycle as

$$B = \frac{b_0}{2} + \sum(b_k\cos kpt + a_k\sin kpt), \tag{22}$$

where

$$\begin{aligned} a_k &= \frac{1}{\pi}\int_0^{2\pi} f(pt)\sin kpt\, d(pt), \\ b_k &= \frac{1}{\pi}\int_0^{2\pi} f(pt)\cos kpt\, d(pt). \end{aligned} \tag{22a}$$

For our particular case we have, since $B = B_1$ for the first half of the cycle, and $B = B_2$ for the second:

$$\pi a_k = \int_{-\pi}^{0} B_2(h, H)\sin kpt\, d(pt) + \int_0^{\pi} B_1(h, H)\sin kpt\, d(pt),$$

$$\pi b_k = \int_{-\pi}^{0} B_2(h, H)\cos kpt\, d(pt) + \int_0^{\pi} B_1(h, H)\cos kpt\, d(pt).$$

These integrals may be simplified considerably when we take advantage of the fact that both B_1 and B_2 are even functions of the time as given by (21). Thus

$$\pi a_k = \int_0^{\pi} [B_1(h, H) - B_2(h, H)]\sin kpt\, d(pt),$$

$$\pi b_k = \int_0^{\pi} [B_1(h, H) + B_2(h, H)]\cos kpt\, d(pt).$$

Referring to (21) we may then write

$$\begin{aligned} a_k &= \frac{2}{\pi}\int_0^{\pi} (A + C\cos 2pt)\sin kpt\, d(pt), \\ b_k &= \frac{2}{\pi}\int_0^{\pi} (B\cos pt + D\cos 3pt)\cos kpt\, d(pt). \end{aligned} \tag{23}$$

[17] This coefficient is not to be confounded with the general expression for flux density.

Upon integration of (23) the coefficients for the fundamental and third harmonic flux components are found to be as follows:

$$\begin{aligned} a_1 &= \frac{4}{\pi}(A - C/3), & b_1 &= B, \\ a_3 &= \frac{4}{\pi}\left(\frac{A}{3} + \frac{3C}{5}\right), & b_3 &= D, \end{aligned} \tag{24}$$

which, by reference to (20), may be put in terms of the branch coefficients.

Appendix 4—Impedance Reaction to a Small Third Harmonic in the Presence of a Large Fundamental

We have for the two hysteresis branch equations from Eqs. (4*a*), (5*a*)

$$\begin{aligned} B_1(h, H) &= \quad B(O, H) + \beta h + \gamma h^2 + a_{30}h^3 + \cdots, \\ B_2(h, H) &= -B(O, H) + \beta h - \gamma h^2 + a_{30}h^3 + \cdots, \end{aligned} \tag{30}$$

in which

$$\begin{aligned} \beta &= a_{10} + a_{11}H + a_{12}H^2, \\ \gamma &= -(a_{02} + a_{03}H). \end{aligned} \tag{31}$$

Putting (29) in (30) we get

$$\begin{aligned} B(h, H) = A &+ B\cos pt + C\cos 2pt + D\cos 3pt + F\cos npt \\ &+ G[\cos (n+1)pt + \cos (n-1)pt] \\ &+ J[\cos (n+2)pt + \cos (n-2)pt] \end{aligned} \tag{32}$$

in which the coefficients have the following significance

$$\begin{aligned} A &= B(O, H) + H_1^2\gamma/2, & F &= \beta H_3 + 3a_{30}H_1^2H_3/2, \\ B &= \beta H_1 + 3a_{30}H_1^3/4, & G &= \gamma H_1 H_3, \\ C &= \gamma H_1^2/2, & J &= 3a_{30}H_1^2H_3/4. \\ D &= a_{30}H_1^3/4, \end{aligned} \tag{33}$$

The coefficients of the Fourier Series for the output wave may now be obtained as before by combining the two equations (30) since each one is operative during one-half the cycle. There results an expression similar to the one obtained in the single frequency case, and since we have

$$I = b_0/2 + \Sigma a_k \sin kpt + \Sigma b_k \cos kpt$$

the coefficients are evaluated from the expressions

$$a_k = \frac{2}{\pi}\int_0^{2\pi} (A + C\cos 2pt + G(\cos 4pt + \cos 2pt))\sin kpt\, d(pt),$$
$$b_k = \frac{2}{\pi}\int_0^{2\pi} (B\cos pt + (D + F)\cos 3pt + J(\cos 5pt + \cos pt))\cos kpt\, d(pt). \tag{34}$$

Upon integration we find

$$b_1 = B, \qquad b_3 = F. \tag{35}$$

Comparing these two coefficients we see that at low amplitudes we may write

$$b_1 = \mu H_1, \qquad b_3 = \mu H_3,$$

in which the permeability is the same to the two components, and is determined by the fundamental amplitude. For the dissipative terms we find

$$a_1 = \frac{4}{\pi}(A - C/3 - 2G/5),$$
$$a_3 = \frac{4}{\pi}(A/3 - 3C/5 + 6G/35), \tag{35'}$$

but some care is required in interpreting these expressions. Inasmuch as we are primarily interested here in determining the dissipative component to a third harmonic magnetizing force of amplitude H_3, we are required to select from a_3 only those terms containing H_3, which means the single term $24G/35\pi$. The other terms take care of the harmonic producing properties of the core and do not affect the impedance to the third harmonic. The impedance term for the third harmonic comes down to

$$\frac{24}{35\pi}(a_{02}H_1H_3 + a_{03}H_1^2H_3),$$

which may be written as

$$\frac{24}{35\pi}H_3\frac{B(O, H)}{H_1}.$$

This may be compared with the corresponding term for the fundamental given by (25).

Appendix 5. Effect of Air-Gap by Vacuum Tube Analogy

In the elementary treatment of non-linear two element vacuum tube circuits, approximate solutions are obtained in the form

$$J = J_1 + J_2 + \cdots J_n,$$

where J represents the variable part of the space current on the basis that J_n represents the nth approximation, but that the series converges rapidly so that we need consider only the first two terms to arrive at a substantially accurate result. The expressions derived for the first and second approximations are known to be

$$J_1 = a_1E/(1 + a_1R),$$
$$J_2 = a_2E^2/(1 + a_1R)^3,$$

where the 'a' coefficients describe the tube characteristic

$$J = a_1v + a_2v^2,$$

R being the external plate circuit resistance, v the tube potential, and E the circuit e.m.f.

Turning now to the equations for the hysteresis loop branches, we have from (4a)

$$B = a_{10}h - a_{02}h^2 + a_{11}hH + a_{02}H^2.$$

Hence by the analogy between flux and current, and between reluctance and resistance, the first order terms are reduced by the factor $1/(1 + a_1R)$ which corresponds to $1/(1 + \lambda a_{10}/A)$, and the second order terms are reduced by the cube of this factor, which yields the same results (42) as the laborious direct method.

Airways Communication Service [1]

By EDWARD B. CRAFT

THE present development of air transport is bringing out its need for adequate communication in much the same manner as the earlier development of railway operations disclosed for that industry the necessity of special communication services if speed and density of traffic were to be obtained with safety. The electric telegraph by a most fortunate coincidence was available just at the time the railways required it; and as the demand for speed became pressing the telephone was perfected. Today the railways of the country, in general, use the telegraph for administrative messages, where a written record is wanted, and use the telephone for despatching, where speed and accuracy are primary requirements.

By another fortunate coincidence, radio appears to be available just at the time it is needed for communication with aircraft in flight. Radio in the form of either telegraph or telephone has been highly developed for communication between points on the surface of the globe. For communication between aircraft and airports it is available in principle although not yet so well developed. During the war, both in this country and abroad, radio equipment of relatively crude design was installed in aircraft and proved of great utility. Since the war, radio telegraphy for aircraft has been further developed by the naval and military services, but radio telephony has received less attention, probably because of the inherent difficulties and lack of a pressing demand.

Following the remarkable success of the Air Mail and the passage of the Air Commerce Act of 1926, we are now fairly launched into an era of air transport of mails, express and passengers. National Airways, laid out and equipped by the Department of Commerce under authority of the Air Commerce Act, already compare in extent with the main trunk line mileage of the railways. Scheduled flying over these airways goes on by night as well as by day. A commercial degree of reliability and safety has been reached in so far as the airplane and its engine are concerned and, when surprises due to bad weather can be eliminated, the safety of air transport should compare favorably with that of other forms of transportation.

Although weather is beyond our control, meteorological science is able to forecast its major phenomena with a high degree of precision,

[1] Contributed to *Aviation* for October 1928.

provided data describing present and past weather conditions can be collected from a sufficient number of places. The progress of a weather disturbance can be tracked and the time of its arrival at a given point predicted. By means of a suitable communication system weather reports from observers located along and near an airway can be collected; and it should be possible, therefore, to reduce materially the weather hazard of air transport.

A full-scale meteorological experiment of this nature is now being conducted in California by the Weather Bureau with the cooperation of the Guggenheim Fund for the Promotion of Aeronautics and of the Pacific Telephone and Telegraph Company. Meteorologists at the Oakland and Los Angeles airports receive several times a day, by long distance telephone, weather data from observers at a large number of selected points in the state. After an exchange of these collected data, these meteorologists forecast flying-weather for aviators starting out over the airway between these airports. The experiment will be continued until the value of the special weather service can be estimated.

Since the communications problem of safe air transport presented features which in a number of respects were unique, it was referred by the Interdepartmental Committee on Aeronautical Meteorology to experts of the American Telephone and Telegraph Company and Bell Telephone Laboratories. What was desired was the collection of reports from a considerable number of widely distributed observers in a relatively short interval of time, say, from twenty observers in twenty minutes. Naturally, it is not commercially practicable to call the party desired, set up the connections, have him answer and give his data all in the space of one minute. However, an equivalent result has been obtained by evolving a special telephone procedure for the purpose. At the appointed time a team of long-distance telephone operators call up successively the listed observers. Each as he answers is asked to hold the line and wait his turn when the operator connects him to the airport meteorologist.

It has been found by trial that the weather data can be reported and recorded in thirty seconds. Consequently, the list of observers can be readily gone through if one minute each is allowed. To the Los Angeles and Oakland airports about forty observers are now reporting weather five times a day. These collected reports are exchanged between airports; and airplanes starting over the airway are provided with a forecast of the weather they may expect enroute and upon arrival.

On the basis of these forecasts, it is hoped that the pilots may be able to avoid bad weather by choosing an alternative route or

by selecting the terminal field where weather conditions are more propitious. Both Los Angeles and the San Francisco Bay region have several airports and there are two routes between them, one up the valley via Bakersfield, and the other the more direct line to the west. The experiment will be carried on for a full year and so cover the complete cycle of the seasons. On the basis of the demonstrated value of this service to the users of the airway, the matter of its continuance or possible extension to other airways can then be decided by the Weather Bureau. Unfortunately, however, California weather is proverbially good, and the experiment will, therefore, be concerned mainly with local fog and visibility conditions. It is possible also that interests other than aeronautical may discover advantages in a short range forecast of local weather. If so, the value of the experiment will be correspondingly increased.

Weather data are also being collected in the east from observers in New Jersey and Pennsylvania by the meteorologist at Hadley Field who employs a somewhat similar method of sequence operation of the long-distance telephone lines.

In addition to the problem of collecting weather data, there is the closely related matter of distributing local weather reports and forecasts between airports. This is "point-to-point service." It may be accomplished by a special radio-telegraph network, by commercial telegraph or by long-distance telephone, and over private or leased wires either by telephone or by telegraph. Local conditions, volume of traffic and economic considerations, in general, determine which type of service should be provided.

Besides its use for weather messages, point-to-point communication between landing fields along an airway is desirable for following the progress of an airplane with its passengers and cargo. Such a despatching service is somewhat analogous to that of a railway and is a necessity if scheduled connections with trains and other aircraft are to be met. Also, there is the necessity of accountability for mails and express; for example, on departure the landing fields ahead must be informed not only of the fact of starting but of what mail is on board. Upon landing there must be a message announcing the event. In this way the progress of a plane can be followed by the terminal airports.

Although air transport of passengers has not yet reached a large volume in this country, European experience indicates that we soon will be concerned with communication problems having to do with passengers' convenience and comfort. Train and bus connections, hotel accommodations and meals, will have to be arranged for by the traffic department of an air transport company.

Point-to-point communication facilities are also required for the general administrative business of the airway and of the air transport companies.

Along our present airways at short intervals are intermediate landing fields upon which planes may land when forced down by weather or mechanical trouble. Such landings, however, are infrequent and will presumably become increasingly rare; but when a forced landing does take place instant communication with the nearest airport is urgent on account of passengers, mail, and the air transport company itself. Telephones are now provided at these intermediate fields by the Department of Commerce and kept available for such emergency use. The same telephones can be used, of course, for the routine collection of weather data by the airport meteorologist.

On some airways communication between terminal landing fields or airports is now handled by radio telegraph and on others by long-distance telephone. Neither system is ideal for the purpose. Radio telegraph is slow and is often unreliable in times of bad static when weather messages become urgent. It also utilizes radio ether channels which are needed for communication with planes. Moreover, a telegraph operator must be constantly listening throughout the twenty-four hours even though messages come infrequently. Commercial wire telephone service on the other hand although generally fast and reliable provides no written record of the messages, nor does it economically repeat messages at such other and distant airports as may be interested. Weather conditions at Cleveland, for example, are of interest both to New York and to Chicago airports. Likewise the time of departure of the New York air mail from Chicago is of interest to all landing fields enroute.

An ideal system which is instantaneous and reliable, repeats messages at all airports, is free from interference, takes up no radio channels, and furnishes a permanent record of all messages at all airports, is the telephone-typewriter service. Telephone-typewriter systems make possible the instantaneous transmission of communications between distant offices and provide simultaneously each office and any desired intermediate stations with typewritten copies. This service has been used for a good many years by the principal press associations and is now being extended rapidly to serve the needs of our larger business organizations.

To utilize the telephone-typewriter system along an airway requires only the installation of keyboard transmitting apparatus and tape printing apparatus at terminal fields and their interconnection by a private or leased wire circuit. Then anyone familiar with a type-

writer may type a message which will appear on the tape fed automatically from the apparatus at every other connected point. The message is automatically and permanently recorded under the control of the sending station. Constant attendance or listening-in is, therefore, not required; and operators at the various receiving points are thus free to attend to telephone calls from intermediate fields, to operate radio beacons and lights, and to carry on whatever duties may be assigned to them.

Telephone-typewriter service has been initiated by the Department of Commerce at Hadley Field, at Cleveland, at Chicago and at San Francisco, where in each place the local radio stations, weather bureau offices and the airport offices are all interconnected. It is planned, at a later date, to equip experimentally some airway with complete telephone-typewriter service between airports.

When an aviator leaves an airport he should be given information of the weather along the route ahead of him and a forecast of the nature of probable changes during the time of his flight. If general weather conditions are settled, or if his flight is a short one, a forecast is entirely adequate. However, for long flights and at times of uncertain and threatening weather, it is important that the pilot be continuously advised by radio of the weather conditions he may encounter during his flight. In particular, reports of the visibility and landing conditions at the airport where he expects to land and storm warnings should be sent him. Weather and landing advice can be broadcast from each airport along the airway. Provision of radio transmitters at airports and receiving sets in the planes will make possible a simple one-way system of communication and will permit any number of planes in the air to be advised without confusion.

The Department of Commerce, in its program of Aids for Air Navigation plans to install radio-telephone transmitters at principal terminal fields to broadcast, to planes in flight, weather and landing information. In addition, there will be a radio-beacon service to assist pilots in finding the landing field.

European practice, however, has not developed a broadcasting service along this line but has evolved a two-way system in which the pilot of the airplane talks with the nearest airport. Such a system has obvious advantages where it is desired by an air transport company to instruct or control rather than merely inform its aviators. The obvious disadvantage lies in the fact that on a single radio channel the airport can converse with only a single airplane at a time. On the London-Paris airway, it is reported, the practice has recently been adopted of communicating on one channel by radio telegraph with the

large planes which carry a radio operator and on another channel by radio telephone with the smaller planes.

Two-way communication has the great and obvious merit of permitting a pilot to discuss the weather outlook with an airport meteorologist, to consider alternative landing places in view of such factors as his remaining fuel supply or the direction of wind, and to decide if necessary on a change in landing place and to be assured of arrangements there for the care of his passengers and mail. It seems reason-



FIG. 1. THE WHIPPANY RADIO LABORATORY.

able, therefore, to predict that operators of air transport fleets will require two-way communication with their planes in flight, although taxi services and private owners without ground organization along the airway may, in general, be content with a public one-way broadcasting service.

Whether one-way or two-way communication is desired for plane-to-ground use it appears that radio telephony as distinguished from telegraphy will be essential. Radio telegraphy requires on board the plane the individual attention of a special radio operator for sending and receiving. Although very large multi-engined passenger planes will certainly carry a relief pilot in flight, it is doubtful whether good

commercial pilots can be made into good telegraphers and vice versa. For long distance over-sea flights and for expeditionary purposes the radio telegraph has, without doubt, preponderating advantages of longer range with the same transmitter power and of intelligibility through a higher level of interfering signals and acoustic noise on board, aside from its convenience in communication enroute with surface vessels. For regular service on established airways, however, the telephone is undoubtedly superior.

The perfection of facilities for communicating weather and landing information to planes in flight, which will enable them to operate with safety under relatively unfavorable meteorological conditions, will greatly stimulate the demand for improved aids to navigation. It seems to be established that flying under conditions of poor visibility, when landmarks are totally obscured and beacon lights are useless, requires some form of radio goniometry if the pilot is to find his way through.

A number of systems have been proposed for this purpose. The London-Paris Airway is equipped with radio direction-finding equipment on the ground by means of which the position of planes can be determined on request. The disadvantages of this arrangement lie mainly in its relative slowness and its lack of traffic capacity. The radio beacon of the type being developed by the Bureau of Standards, giving an equi-signal zone which can be observed by the plane, is free from these objections. It is, however, subject to the disadvantage that it indicates a straightline course which cannot always coincide with the airway and is of little value if detours are required to avoid storm centers and foggy areas.

Another system, a recent development of the British Royal Air Force, employs a rotating loop transmitter at the ground station and indicates the bearing of the plane with respect to the transmitter by means of a special stop watch. This system is relatively slow but permits the pilot to navigate as he would if one or more beacon lights were visible. All of these various methods of goniometry have special advantages and disadvantages, and occupy more or less of the valuable and restricted ether space. The evolution of the system which is most satisfactory will be a matter of time and will require close co-operation on the part of all factors in the industry.

Bell Telephone Laboratories, at its radio station at Whippany, New Jersey, has erected an experimental two-way radio-telephone system and radio beacon. In connection with this apparatus it utilizes a Fairchild Cabin Monoplane with Pratt and Whitney wasp engine. The plane has been carefully bonded and shielded and is

equipped with radio field-measuring apparatus of the Laboratories' design. With this plane exact measurements can be made at various altitudes under different weather conditions of the efficiency of radio transmission from the Whippany transmitter. In addition the plane carries radio transmitting and receiving sets of experimental design.



FIG. 2. THE CABIN MONOPLANE FOR EXPERIMENT IN AIRWAYS COMMUNICATION.

It is, in fact, a flying radio laboratory in which the engineers may experiment under actual flying conditions.

Whether a radio beacon service and a radio telephone service at all the various airports over the country can be made practicable is largely a question of available ether channels. By international agreement, the frequency band 285–315 kcs. (1050–950 meters) is reserved for radio beacons, both marine and air service. For "air mobile service exclusively" there is reserved the band 315–350 kcs. (950–850 meters) in which the 900 meter wave (333 kcs.) is reserved as an air service calling wave and is not to be assigned.

Radio telephony requires a band of frequencies sufficiently wide to include the "side bands" of speech frequency. For distinct transmission of speech, neglecting certain requirements of musical quality, this might require a minimum of 6,000 cycles. In this band reserved for "air mobile service exclusively" there is room, therefore, for but

three telephone channels above and three below the calling wave, or a total of six channels. Assuming that a beacon requires a channel width of but 300 cycles, there are altogether for marine and airport use one hundred beacon channels in the band 285–315 kcs.



FIG. 3. CABIN LABORATORIES OF THE MONOPLANE.

The band reserved for beacons is already partly occupied by marine beacons, and near the coast difficulty may arise in finding clear channels for airport beacons. Although it is probable that, by a proper geographical distribution of frequencies, there may be worked out without

undue interference an adequate beacon service we can make no assumption that any extra space can be found in the beacon band for radio telephony.

A radio telephone system with a sufficiently powerful transmitter and sufficiently sensitive receiver to give reliable communication for 100 miles will give fair communication for perhaps 200 miles, and its carrier wave will interfere with reception for a much greater distance. To avoid interference due to the beating of carrier frequencies, airports within a few hundred miles of one another may be assigned to different frequency channels, but serious difficulty is at once apparent from a map of the National Airways. Within 800 miles of Chicago, for example, there are over fifty terminal fields or airports. It would seem obviously impractical to assign the available six telephone channels to cover the eastern and central United States without serious interference. By restricting power as much as possible and by other means yet to be devised, it may be found possible to assign the same wave-length to airports relatively nearer together. For the distribution of weather information only, however, the airways may well find insufficient the frequencies in the exclusive band, 315–350 kilocycles.

On certain main routes, air transport companies will eventually require two-way telephone despatching systems of their own to control plane movements. These systems will consist of radio stations situated at the various airports along the route and interconnected by suitable wire lines. The frequency channels required for such services cannot be found in the 315–350 kilocycles band which, as just indicated, is apparently inadequate for the public services of weather broadcasting from airports. Further channels in the short-wave region appear to be necessary.

In the short wave region Bell Telephone Laboratories have initiated an additional development project. In cooperation with the Boeing Air Transport Company, the Laboratories have undertaken to survey the Chicago-San Francisco Airway and to develop a system of two-way telephony between planes in flight and terminal landing fields on this route. The Boeing Company planes and landing fields will be equipped with experimental radio apparatus and a joint full-scale experiment will be conducted during the winter of 1928–29. From this work it is hoped to determine for an air transport company the requirements for a two-way radio telephone service. The investigation will furnish the basis for offering such facilities to other air transport operators.

This development of two-way radio-telephony on short waves is

entirely distinct from the government's program of Aids to Air Navigation. That service contemplates one-way radio telephony and direction finding on long waves. The government service is to be available to all flyers who equip themselves to receive it. The two-way system is for private communication and despatching service of air transport companies which wish to control their planes in flight, and to remain in constant communication with their pilots and passengers.

Also, although not yet required, it can safely be predicted that at busy airports there will soon arise a need for radio means to control precedence in the take-off and landing of airplanes. This virtually amounts to traffic control and can be accomplished by low-power two-way radio telephone. Planes wishing to land may announce themselves and remain aloft until directed by the airport manager in the control tower to land at a designated part of the field.

In all these present and future problems, it is the policy of the American Telephone and Telegraph Company and the Bell System to assist by developing ways and means for making available to commercial aviation the best possible communication service.

Abstracts of Bell System Technical Papers Not Appearing in this Journal

Influence of Carbon and Silicon Variations in Grey Cast Iron.[1] D. G. ANDERSON and G. R. BESSMER. In this short article the author gives the results of a series of tests of grey cast irons with different carbon and silicon contents. Three series were run in each of which the silicon content was kept constant and the amount of carbon varied. The results indicated that with two percent silicon the carbon content may be reduced without materially increasing the amount of combined carbon. This results in some improvement in the physical properties of the iron.

Strength-Tests of Telephone Materials.[2] J. R. TOWNSEND. Static tests, such as the ordinary tension or torsion tests, have fallen somewhat into disrepute during the last ten years, the author claims, as the ultimate strengths obtained from them are not always indicative of the forces materials will withstand in actual service. Their place is being taken by repeated-stress tests in which the sample is subjected to conditions more nearly representing those met in ordinary service. In illustration the author mentions several tests of this class being applied in Bell Telephone Laboratories on cable sheath material and springs.

The Reduction of Atmospheric Disturbances.[3] JOHN R. CARSON. In the decade or so during which the problem of eliminating or at least reducing atmospheric disturbances has been given serious and systematic study we have learned, more or less definitely, what we can and cannot do in this direction. For example, we know that there are definite limits to what can be accomplished by frequency selection. We know that directional selectivity is of substantial value, particularly when the predominant interference comes from a direction other than that of the desired signal, and we can calculate pretty well the gain to be expected from a given design.

The object of this note is to analyze another arrangement which provides for high-frequency selection plus low-frequency balancing after detection. The broad idea of balancing out the interference is old, but no general analysis of the arrangement seems to have been made. Furthermore the principle of balance has recently acquired

[1] "Fuels and Furnaces," Vol. VI, No. 7, pp. 957 and 972, July, 1928.

[2] "Instruments," Vol. 1, No. 7, pp. 313–315, July, 1928.

[3] *Proceedings of the Institute of Radio Engineers*, July, 1928, Vol. 16, No. 7, pp. 966–975.

fresh interest due to the system disclosed by Armstrong [4] in which high-frequency selectivity and low-frequency balancing are essential features. Armstrong's scheme is treated in more detail in the latter part of this paper.

The conclusions of this study are entirely negative, that is, no appreciable gain is to be expected from balancing arrangements. This is quite in agreement with the conclusion drawn over ten years ago as a result of a rather extended experimental study made in the Bell System. In fact, as more and more schemes are analyzed and tested, and as the essential nature of the problem is more clearly perceived, we are unavoidably forced to the conclusion that static, like the poor, will always be with us.

Thermostat Design for Frequency Standards.[5] W. A. MARRISON. A means for maintaining constant temperature is described in which those temperature variations which are essential for operation of the controlling element are prevented from reaching the controlled chamber by a wall of material especially chosen for the purpose. Such a wall 1 cm. thick, consisting of alternate thin layers of felt and copper, will reduce temperature variations having a period of one minute or less by a factor of 10,000 to 1.

Technical Considerations Involved in the Allocation of Short Waves; Frequencies between 1.5 and 30 Megacycles.[6] LLOYD ESPENSCHIED. This short paper discusses the relation between frequency and distance of transmission for short waves in so far as it affects allocation. A table is given in which the entire short-wave field from 10 to 200 meters is divided into three major bands each containing numerous sub-bands. For each sub-band the number of channels theoretically possible is given and also the number of channels being used at the present time. Factors affecting the separation of channels are also listed.

Effect of Street Railway Mercury Arc Rectifiers on Communication Circuits.[7] CHARLES J. DALY. This paper describes the effects experienced on the telephone circuits from two mercury arc rectifier substations recently installed in Bridgeport, Conn., and shows in table form the relative magnitude of the interfering effects between rotating equipment and mercury arc rectifiers as a means of energizing the street railway system. The method and the type of apparatus used to reduce the effects experienced from the rectifiers are also described.

[4] *Proceedings of the Institute of Radio Engineers*, Jan., 1928, Vol. 16, No. 1, p. 15.
[5] *Proceedings of the I. R. E.*, Vol. 16, No. 7, pp. 976–980, July, 1928.
[6] *Proceedings of the I. R. E.*, Vol. 16, No. 6, pp. 773–777, June, 1928.
[7] *Journal of the A. I. E. E.*, Vol. XLVII, No. 7, pp. 503–506, July, 1928.

Compressed Powdered Permalloy—Manufacture and Magnetic Properties.[8] W. J. SHACKELTON and I. G. BARBER. The paper gives a brief description of the manufacture of magnetic cores of compressed permalloy powder followed by information covering their magnetic properties with particular reference to their use in loading coils. Production of the powder, and its insulation, pressing and annealing, are discussed. Under magnetic properties, permeability, core loss, and modulation are treated. Curves are given illustrating the characteristics of interest in connection with the design and application of loading coils; and comparisons to corresponding characteristics of compressed powdered iron are made throughout.

Thermal Agitation of Electric Charge in Conductors.[9] H. NYQUIST. The electromotive force due to thermal agitation in conductors is calculated by means of principles in thermodynamics and statistical mechanics. The results obtained agree with results obtained experimentally.

Time-Lag in Magnetization.[10] RICHARD M. BOZORTH. An investigation has been made of the time-lag in magnetization in a permalloy wire to determine whether lag can be satisfactorily accounted for as due to eddy-currents alone or whether permalloy shows a marked magnetic viscosity such as has been observed by Ewing in iron wires. Eddy-current lag has been calculated approximately in a manner which takes into account the changing slope of the magnetization curve. A comparison of the calculated and observed magnetization-*vs.*-time curves indicates that the effect is well accounted for as eddy-current lag alone. The eddy-current lag has also been calculated for an iron ring, for which the time-lag has been reported recently in a number of papers by Lapp. The time-lag which he observed is satisfactorily accounted for as eddy-current lag instead of as magnetic viscosity as he had supposed.

Thermal Agitation of Electricity in Conductors.[11] J. B. JOHNSON. Statistical fluctuation of electric charge exists in all conductors, producing random variation of potential between the ends of the conductor. The effect of these fluctuations has been measured by a vacuum tube amplifier and thermocouple, and can be expressed by the formula $\bar{I}^2 = (2kT/\pi)\int_0^\infty R(\omega)\,|\,Y(\omega)\,|^2 d\omega$. I is the observed current in the thermocouple, k is Boltzmann's gas constant, T is the absolute

[8] *Journal of the A. I. E. E.*, Vol. XLII, No. 6, pp. 437–440, June, 1928.
[9] *Physical Review*, Vol. 32, No. 1, pp. 110–113, July, 1928.
[10] *Physical Review*, Vol. 32, No. 1, pp. 124–132, July, 1928.
[11] *Physical Review*, Vol. 32, No. 1, pp. 97–109, July, 1928.

temperature of the conductor, $R(\omega)$ is the real component of impedance of the conductor, $Y(\omega)$ is the transfer impedance of the amplifier, and $\omega/2\pi = f$ represents frequency. The value of Boltzmann's constant obtained from the measurements lies near the accepted value of this constant. The technical aspects of the disturbance are discussed. In an amplifier having a range of 5,000 cycles and the input resistance R, the power equivalent of the effect is $\overline{V}^2/R = 0.8 \times 10^{-16}$ watt, with corresponding power for other ranges of frequency. The least contribution of tube noise is equivalent to that of a resistance $R_e = 1.5 \times 10^5 i_p/\mu$, where i_p is the space current in milliamperes and μ is the effective amplification of the tube.

The Voltage-Current Relation in Central Cathode Photoelectric Cells.[12] Thornton C. Fry and Herbert E. Ives. This paper presents a theoretical basis for the interpretation of the experimental results described in the paper which follows. It considers a source of photoelectrons located on the inner of two concentric spheres; derives the trajectory of an electron shot off at any angle with any speed; and then makes use of this information to compute the current which would be received by a small collector located anywhere on the outer sphere upon very general assumptions as to the directional distribution and velocity distribution of the photoelectrons. This theoretical study is followed by graphical presentation of results computed for several typical cases of special interest in connection with the experimental study.

The Distribution in Direction of Photoelectrons from Alkali Metal Surfaces.[13] Hervert E. Ives, A. R. Olpin and A. L. Johnsrud. An experimental study of the distribution in direction of photoelectrons emitted from alkali metal surfaces irradiated by light incident at various angles and polarized in different planes. The alkali metal surfaces used were of two sorts: (1) liquid alloys of sodium and potassium, (2) thin films of potassium or rubidium on polished platinum. In all cases the alkali metal surface was at the center of a large spherical enclosing anode, provided either with collecting tabs at various angular positions or with an exploring finger. It is found that the emission closely obeys Lambert's law, but that the ellipse by which the emission is represented, in polar coordinates, is more elongated normally to the surface for perpendicularly incident light than for obliquely, when the direction of the electric vector is in both cases parallel to the surface, and still more elongated for obliquely incident light with the

[12] *Physical Review*, Vol. 32, No. 1, pp. 44–56, July, 1928.
[13] *Physical Review*, Vol. 32, No. 1, pp. 57–80, July, 1928.

electric vector in the plane of incidence. The distribution curves are all perfectly symmetrical about the normal to the surface, showing no tendency to follow the direction of the electric vector.

Oscillographic Observations on the Direction of Propagation and Fading of Short Waves.[14] H. T. FRIIS. The short-wave transmission path is generally but not always located in the vertical plane through the transmission and receiving points.

Direction finding depends upon determining the direction of the wave at the receiving point; it does not give accurate results when the twilight zone is in the way of the wave path.

The angle between the earth and the direction of short-wave propagation varies continuously and the changes in this angle are much larger than the changes in angle of propagation in the horizontal plane.

The observations are consistent with the view that the fading is mainly caused by wave interference.

An Improved Permeameter for Testing Magnet Steel.[15] B. J. BABBITT. The increasing use of cobalt steel in the manufacture of permanent magnets has created a need for a permeameter that is capable of determining accurately the magnetic properties of such steel in bar form. The common commercial permeameters are not capable of producing the high magnetizing forces required for this purpose. Commercial permeameters are chiefly of two types, the yoke type and the Burrows type. The latter is difficult to operate and requires an experienced operator for a reasonable output; it cannot be adapted to the testing of cobalt steel unless it is practically rebuilt throughout. The yoke type of permeameter may be adapted to the testing of cobalt steel by the use of extensions to the poles so that the distance between them is much less. In this way the greater part of the magnetomotive force is distributed over a short portion of the magnetic circuit and the magnetomotive force per centimeter is correspondingly greater. The permeameter that is described below has been developed by the Magnetic Materials Division at the Hawthorne Works of the Western Electric Company to overcome the chief objections common to present commercial permeameters.

Corrosion of Cable Sheath in Creosoted Wood Conduit.[16] R. M. BURNS and B. A. FREED. This paper deals with the identification of a corrosion of lead cable placed in creosoted wood conduit, and with the

[14] *Proceedings of the Institute of Radio Engineers*, May, 1928, Vol. 16, No. 5, pp, 658–665.

[15] *Journal of the Optical Society of America and Review of Scientific Instruments* Vol. 17, No. 1, pp. 47–58, July, 1928.

[16] *Journal of the A. I. E. E.*, Vol. XLVII, No. 8, pp. 576–579, August, 1928.

determination and application of methods of allaying it. The trouble was experienced mainly on the Pacific Coast where, although Douglas fir conduit was introduced about 1911, the first case of corrosion which could definitely be ascribed to the creosoted conduit did not occur till 1921.

A search for the cause of the trouble led to making systematic analyses of the air present in the conduit and these analyses revealed the presence of acetic acid in sufficient amount to account for the corrosion in the presence of carbon dioxide which was also shown to be present.

After much experimenting a method was developed to stop the corrosion by pumping ammonia into the ducts. Results have been very satisfactory and seem to indicate that a single treatment is sufficient.

Small Samples—New Experimental Results.[17] W. A. SHEWHART and F. W. WINTERS. This article reviews briefly the Theory of Errors of Averages, paying particular attention to some of the most recent work in connection with small samples. New empirical results are presented showing the advantage that arises from the use of the latest error theory and pointing out the effect of the limitations imposed upon it. The information contained in this paper indicates that further theoretical studies are necessary in order that the application of small sample theory may give more accurate solutions to the problems that arise in practice.

[17] *Journal of American Statistical Association*, New Series, No. 162 (Vol. XXIII), pp. 144–153, June, 1928.

Contributors to this Issue

George A. Campbell, B.S., Massachusetts Institute of Technology, 1891; A.B., Harvard, 1892; Ph.D., 1901; Gottingen, Vienna and Paris, 1893–96; Mechanical Department, American Bell Telephone Company, 1897; Engineering Department, American Telephone and Telegraph Company, 1903–19; Department of Development and Research, 1919–; Research Engineer, 1908–. Dr. Campbell has published papers on loading and the theory of electric circuits, including electric wave-filters, and is also well known to telephone engineers for his contributions to repeater and substation circuits.

C. W. Robbins, Eureka Electric Company, 1901–1905; Western Electric Company; Testing Apparatus Design, 1905–06; Chief of Inspection Methods Division, Chicago, 1906–08; Chief Inspector Cable, Rubber and Insulating Shops, Hawthorne, 1908–18; Assistant Superintendent of Inspection, 1918–27; Assistant Superintendent of Inspection Development, 1927–. Much of Mr. Robbins' work has been connected with gages, testing and measuring apparatus and methods.

Karl K. Darrow, S.B., University of Chicago, 1911, University of Paris, 1911–12, University of Berlin, 1912; Ph.D. in physics and mathematics, University of Chicago, 1917; Engineering Department, Western Electric Company, 1917–24; Bell Telephone Laboratories, Inc., 1925–. Mr. Darrow has been engaged largely in preparing studies and analyses of published research in various fields of physics. His earlier articles on Contemporary Physics form the nucleus of a recently published book entitled "Introduction to Contemporary Physics" (D. Van Nostrand Company).

E. Peterson, Cornell University, 1911–14; Brooklyn Polytechnic, E.E., 1917; Columbia, A.M., 1923; Ph.D., 1926; Electrical Testing Laboratories, 1915–17; Signal Corps, U. S. Army, 1917–19; Bell Telephone Laboratories, 1919–. Mr. Peterson's work has been largely in theoretical studies of carrier current apparatus.

Edward B. Craft, Engineering Department, Western Electric Company, Chicago, 1902–07; Development Engineer, Western Electric Company at New York, 1907–18; Assistant Chief Engineer, 1918–22; Chief Engineer, 1922–25; Executive Vice President of Bell Telephone Laboratories, 1925–.

Mr. Craft's duties have been executive for some years, but he also has many patents and has made outstanding individual contributions, prominent among which is the flat type relay.

Index to Volume VII

A

Absorption Coefficient of Porous Materials, Measurement of Acoustic Impedance and the, *E. C. Wente* and *E. H. Bedell*, page 1.

Acoustic Impedance and the Absorption Coefficient of Porous Materials, Measurement of, *E. C. Wente* and *E. H. Bedell*, page 1.

Advances in Physics—XV, Contemporary. The Classical Theory of Light, *K. K. Darrow*, First part, page 281; Second part, page 730.

Affel, H. A., C. S. Demarest and *C. W. Green*, Carrier Systems on Long Distance Telephone Lines, page 564.

Airways communication Service, *E. B. Craft*, page 797.

American Institute of Electrical Engineers, Joint Meeting of the Institution of Electrical Engineers and the, page 161.

Apparatus, Precision Tool-Making for the Manufacture of Telephone, *J. H. Kasley* and *F. P. Hutchison*, page 375.

Automatic Machine Gauging, *C. W. Robbins*, page 708.

B

Bartlett, B. W. and *J. G. Ferguson*, Measurement of Capacitance in Terms of Resistance and Frequency, page 420.

Bedell, E. H. and *E. C. Wente*, Measurement of Acoustic Impedance and the Absorption Coefficient of Porous Materials, page 1.

Blackwell, O. B., Transatlantic Telephony—the Technical Problem, page 168.

Bridge for Measuring Small Time Intervals, *J. Herman*, page 343.

Bridge: Electrical Measurement of Communication Apparatus, *W. J. Shackelton* and *J. G. Ferguson*, page 70.

Bridge: Measurement of Capacitance in Terms of Resistance and Frequency, *J. G. Ferguson* and *B. W. Bartlett*, page 420.

Buckley, O. E., High-Speed Ocean Cable Telegraphy, page 225.

C

Cable, Recent Developments in the Process of Manufacturing Lead-Covered Telephone, *C. D. Hart*, page 321.

Cable Telegraphy, High-Speed Ocean, *O. E. Buckley*, page 225.

Campbell, G. A., Practical Application of the Fourier Integral, page 639.

Capacitance in Terms of Resistance and Frequency, Measurement of, *J. G. Ferguson* and *B. W. Bartlett*, page 420.

Carrier Systems on Long Distance Telephone Lines, *H. A. Affel, C. S. Demarest* and *C. W. Green*, page 564.

Carson, J. R., Present Status of Wire Transmission Theory and Some of its Outstanding Problems, page 268.

Carson, J. R., Rigorous and Approximate Theories of Electrical Transmission along Wires, page 11.

Classical Theory of Light, Contemporary Advances in Physics, XV., *K. K. Darrow*, First part, page 281, Second part, page 730.

Coggins, P. P., Some General Results of Elementary Sampling Theory for Engineering Use, page 26.

Communication Apparatus, Electrical Measurement of, *W. J. Shackelton* and *J. G. Ferguson*, page 70.

Communication Service, Airways, *E. B. Craft*, page 797.

Conductors, Natural Period of Linear, *C. R. Englund*, page 404.

Contemporary Advances in Physics, XV. The Classical Theory of Light, *K. K. Darrow*, First part, page 281, Second part, page 730.

Craft, E. B., Airways Communication Service, page 797.

Crosstalk: Harmonic Production in Ferromagnetic Materials at Low Frequencies and Low Flux Densities, *E. Peterson*, page 762.

Crystal of Nickel, Diffraction of Electrons by a, *C. J. Davisson*, page 90.

D

Darrow, K. K., Contemporary Advances in Physics, XV. The Classical Theory of Light, First part, page 281; Second part, page 730.

Davisson, C. J., Diffraction of Electrons by a Crystal of Nickel, page 90.

Demarest, C. S., H. A. Affel and *C. W. Green*, Carrier Systems on Long Distance Telephone Lines, page 564.

Diffraction of Electrons by a Crystal of Nickel, *C. J. Davisson*, page 90.

Distortion: Harmonic Production in Ferromagnetic Materials at Low Frequencies and Low Flux Densities, *E. Peterson*, page 762.

Distortion and Phase Distortion Correction, Phase, *Sallie Pero Mead*, page 195.

Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks, *O. J. Zobel*, page 438.

Dodge, H. F., Method of Rating Manufactured Product, page 350.

E

Electrical Measurement of Communication Apparatus, *W. J. Shackelton* and *J. G. Ferguson*, page 70.

Electrons by a Crystal of Nickel, Diffraction of, *C. J. Davisson*, page 90.

Englund, C. R., Natural Period of Linear Conductors, page 404.

Equalizers: Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks, *O. J. Zobel*, page 438.

F

Ferguson, J. G. and *B. W. Bartlett*, Measurement of Capacitance in Terms of Resistance and Frequency, page 420.

Ferguson, J. G. and *W. J. Shackelton*, Electrical Measurement of Communication Apparatus, page 70.

Ferromagnetic Materials at Low Frequencies and Low Flux Densities, Harmonic Production in, *E. Peterson*, page 762.

Fourier Integral, Practical Application of the, *G. A. Campbell*, page 639.

Frequency, Measurement of Capacitance in Terms of Resistance and, *J. G. Ferguson* and *B. W. Bartlett*, page 420.

G

Gaging, Automatic Machine, *C. W. Robbins*, page 708.

Green, C. W., H. A. Affel and *C. S. Demarest*, Carrier Systems on Long Distance Telephone Lines, page 564.

Grid Current Modulation, *E. Peterson* and *C. R. Keith*, page 106.

H

Harmonic Production in Ferromagnetic Material at Low Frequencies and Low Flux Densities, *E. Peterson*, page 762.

Hart, C. D., Recent Developments in the Process of Manufacturing Lead-Covered Telephone Cable, page 321.

Hartley, R. V. L., Transmission of Information, page 535.

Herman, J., Bridge for Measuring Small Time Intervals, page 343.

High Efficiency Receiver of Large Power Capacity for Horn-Type Loud Speakers, *E. C. Wente* and *A. L. Thuras*, page 140.

Horn-Type Loud Speakers, High Efficiency Receiver of Large Power Capacity for, *E. C. Wente* and *A. L. Thuras*, page 140.

Hutchison, F. P. and *J. H. Kasley*, Precision Tool-Making for the Manufacture of Telephone Apparatus, page 375.

I

Institution of Electrical Engineers and the American Institute of Electrical Engineers, Joint Meeting of the, page 161.

Integral, Practical Application of the Fourier, *G. A. Campbell*, page 639.

Intervals, Bridge for Measuring Small Time, *J. Herman*, page 343.

J

Joint Meeting of the Institution of Electrical Engineers and the American Institute of Electrical Engineers, page 161.

K

Kasley, J. H. and *F. P. Hutchison*, Precision Tool-Making for the Manufacture of Telephone Apparatus, page 375.

Keith, C. R. and *E. Peterson*, Grid Current Modulation, page 106.

L

Lead-Covered Telephone Cable, Recent Developments in the Process of Manufacturing, *C. D. Hart*, page 321.

Light, Classical Theory of. Contemporary Advances in Physics, XV., *K. K. Darrow*, First part, page 281, Second part, page 730.

Linear Conductors, Natural Period of, *C. R. Englund*, page 404.

Long Distance Telephone Lines, Carrier Systems on, *H. A. Affel, C. S. Demarest* and *C. W. Green*, page 564.

Loud-Speakers, High Efficiency Receiver of Large Power Capacity for Horn-Type. *E. C. Wente* and *A. L. Thuras*, page 140.

M

Machine Gauging, Automatic, *C. W. Robbins*, page 708.

Magnetism: Harmonic Production in Ferromagnetic Materials at Low Frequencies and Low Flux Densities, *E. Peterson*, page 762.

Mead, Sallie Pero, Phase Distortion and Phase Distortion Correction, page 195.

Measurement of Acoustic Impedance and the Absorption Coefficient of Porous Materials, *E. C. Wente* and *E. H. Bedell*, page 1.

Measurement of Capacitance in Terms of Resistance and Frequency, *J. G. Ferguson* and *B. W. Bartlett*, page 420.

Measurement of Communication Apparatus, Electrical, *W. J. Shackelton* and *J. G. Ferguson*, page 70.

Method of Rating Manufactured Product, *H. F. Dodge*, page 350.

Modulation, Grid Current, *E. Peterson* and *C. R. Keith*, page 106.

N

Natural Period of Linear Conductors, *C. R. Englund*, page 404.

Networks, Distortion Correction in Electrical Circuits with Constant Resistance Recurrent, *O. J. Zobel*, page 438.

O

Optics: Contemporary Advances in Physics, XV. The Classical Theory of Light, *K. K. Darrow*, First part, page 281, Second part, page 730.

Oscillations: Natural Period of Linear Conductors, *C. R. Englund*, page 404.

P

Peterson, E., Harmonic Production in Ferromagnetic Materials at Low Frequencies and Low Flux Densities, page 762.

Peterson, E. and *C. R. Keith*, Grid Current Modulation, page 106.

Phase Distortion and Phase Distortion Correction, *Sallie Pero Mead*, page 195.

Physics, Contemporary Advances in, XV. The Classical Theory of Light, *K. K. Darrow*, First part, page 281, Second part, page 730.

Porous Materials, Measurement of Acoustic Impedance and the Absorption Coefficient of, *E. C. Wente* and *E. H. Bedell*, page 1.

Practical Application of the Fourier Integral, *G. A. Campbell*, page 639.

Precision Tool-Making for the Manufacture of Telephone Apparatus, *J. H. Kasley* and *F. P. Hutchison*, page 375.

Present Status of Wire Transmission Theory and Some of its Outstanding Problems, *J. R. Carson*, page 268.

R

Rating Manufactured Product, Method of, *H. F. Dodge*, page 350.

Receiver of Large Power Capacity for Horn-Type Loud Speakers, High Efficiency, *E. C. Wente* and *A. L. Thuras*, page 140.

Recent Developments in the Process of Manufacturing Lead-Covered Telephone Cable, *C. D. Hart*, page 321.

Resistance and Frequency, Measurement of Capacitance in Terms of, *J. G. Ferguson* and *B. W. Bartlett*, page 420.

Rigorous and Approximate Theories of Electrical Transmission along Wires, *J. R. Carson*, page 11.

Robbins, C. W., Automatic Machine Gauging, page 708.

S

Sampling: Method of Rating Manufactured Product, *H. F. Dodge*, page 350.

Sampling Theory for Engineering Use, Some General Results of Elementary, *P. P. Coggins*, page 26.

Shackelton, W. J. and *J. G. Ferguson*, Electrical Measurement of Communication Apparatus, page 70.

Some General Results of Elementary Sampling Theory for Enginereing Use, *P. P. Coggins*, page 26.

Speakers, High Efficiency Receiver of Large Power Capacity for Horn-Type Loud, *E. C. Wente* and *A. L. Thuras*, page 140.

Submarine Cable: High-Speed Ocean Cable Telegraphy, *O. E. Buckley*, page 225.

T

Telegraphy, High-Speed Ocean Cable, *O. E. Buckley*, page 225.

Telephone Apparatus, Precision Tool-Making for the Manufacture of, *J. H. Kasley* and *F. P. Hutchison*, page 375.

Telephony, Transatlantic—the Technical Problem, *O. B. Blackwell*, page 168.

Telephony, Transatlantic—Service and Operating Features, *K. W. Waterson*, page 187.

Theory for Engineering Use, Some General Results of Elementary Sampling, *P. P. Coggins*, page 26.

Thuras, A. L. and *E. C. Wente*, High Efficiency Receiver of Large Power Capacity for Horn-Type Loud Speakers, page 140.

Time Intervals, Bridge for Measuring Small, *J. Herman*, page 343.

Tool-Making for the Manufacture of Telephone Apparatus, Precision, *J. H. Kasley* and *F. P. Hutchison*, page 375.

Transatlantic Telephony—the Technical Problem, *O. B. Blackwell*, page 168.

Transatlantic Telephony—Service and Operating Features, *K. W. Waterson*, page 187.

Transmission along Wires, Rigorous and Approximate Theories of Electrical, *J. R. Carson*, page 11.

Transmission Theory and Some of its Outstanding Problems, Present Status of Wire, *J. R. Carson*, page 268.

Transmission of Information, *R. V. L. Hartley*, page 535.

V

Vacuum Tubes: Grid Current Modulation, *E. Peterson* and *C. R. Keith*, page 106.

W

Waterson, K. W., Transatlantic Telephony—Service and Operating Features, page 187.

Wente, E. C. and *E. H. Bedell*, Measurement of Acoustic Impedance and the Absorption Coefficient of Porous Materials, page 1.

Wente, E. C. and *A. L. Thuras*, High Efficiency Receiver of Large Power Capacity for Horn-Type Loud Speakers, page 140.

Wires, Rigorous and Approximate Theories of Electrical Transmission along, *J. R. Carson*, page 11.

Wire Transmission Theory and Some of its Outstanding Problems, Present Status of, *J. R. Carson*, page 268.

Z

Zobel, O. J., Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks, page 438.