The Bell System Technical Journal

October, 1926

Radio Signaling System for the New York Police Department

By S. E. ANDERSON

SYNOPSIS: By means of the radio signaling system described it is possible, through the addition of a comparatively simple attachment to a standard radio telephone transmitter, to modulate the carrier with an audio frequency tone in such a manner as to provide for calling individually, simultaneously, or in a number of designated groups, any one of several hundred radio receiving stations. At the radio receiving stations apparatus is provided which may be operated from commercial sources of power supply and by means of which a visible or audible signal is given to the operator that a message is about to be broadcast, to which he should listen. Signals are also provided which, in case of improper operation, immediately inform the operator of the points at which attention is required.

INTRODUCTION

FOR some time the New York Police Department has been employing the municipal broadcasting station WNYC to broadcast descriptions of missing persons and other features of police work in which it is desired to enlist the cooperation of the public. The success of this program has been such that the Police Department wished to equip the precinct houses and police booths located in various parts of the city with receiving sets with which they could listen in on communications from the headquarters station WNYC. The fundamental requirement was signaling apparatus incorporated in the receiving sets which would attract the attention of the attendant at the proper times. The system which was finally developed by the engineers of the Bell Telephone Laboratories, Inc., in cooperation with the New York Police Department is an excellent illustration of the adaptability of wire practices to the radio field. The underlying principles employed and much of the apparatus used had previously found extensive application in the Bell System and elsewhere.

The basis of this system is the Western Electric telephone train dispatching system which is in rather general use on railroads throughout the world for the purpose of providing selective signaling on their train dispatching telephone systems. For every division, these systems consist ordinarily of a single line to which are connected a number of stations capable of being called by the dispatcher individually, simultaneously or in groups.[1] This signaling system has also been adapted to radio transmission.[2] Its use permits broadcasting from a central radio transmitting station to police organization districts, patrol boats and automobiles without requiring the constant attention of operators at the receiving stations.

Requirements of the System

Before describing the system which was finally worked out to meet the requirements of the New York Police Department, it seems best to state the nature of these requirements. For a system employed to handle communications ranging all the way from routine messages between police headquarters and its different outlying police stations and patrols to general alarms for insuring the capture of escaping criminals, absolute reliability and flexibility are of the utmost importance. The central station must be able to call the receiving stations individually, collectively, or in a number of designated group combinations corresponding to the police organization districts. To accomplish this result effectively, means must be provided whereby the desired signal may be sent automatically by a simple manual operation. The apparatus for this purpose must be in the form of an attachment which may be used with a standard radio telephone transmitter without extensive modifications.

As the receiver will be in operation continuously, the difficult and expensive maintenance of batteries must be avoided by energizing the vacuum tubes from the commercial power supply system. The tuning arrangements of the receiver must be of the simplest possible character and must be capable of being locked to insure that the receiver remains tuned to the transmitting frequency. The selectivity and sensitivity must be sufficient to insure reliable operation under all conditions. The receiver must provide means for listening to all material broadcast by the central broadcasting station but the signaling system should respond only to signals from the transmitter signaling attachment, irrespective of broadcast speech, music and telegraph signals which may involve the same frequencies. Visible indications should be provided to show when the receiver is in operating condition. The receiver should respond to a call from the central station by operating another visible indicator, in addition to a bell or other audible signal, if desired.

[1] "Modern Methods in Train Dispatching," by J. C. Latham, *Electrical Communication*, Vol. III, No. 1, July, 1924.

[2] "Radio Telephone Signaling Low-Frequency System," by C. S. Demarest, M. L. Almquist and L. M. Clement, *Journal of the A. I. E. E.*, Vol. XLIII, No. 3 March, 1924.

Description of Apparatus

Transmitter Attachment

A photograph of the transmitter attachment is shown in Fig. 1, and a schematic circuit is given in Fig. 2. The apparatus consists of a vacuum tube oscillator and a number of calling keys. These

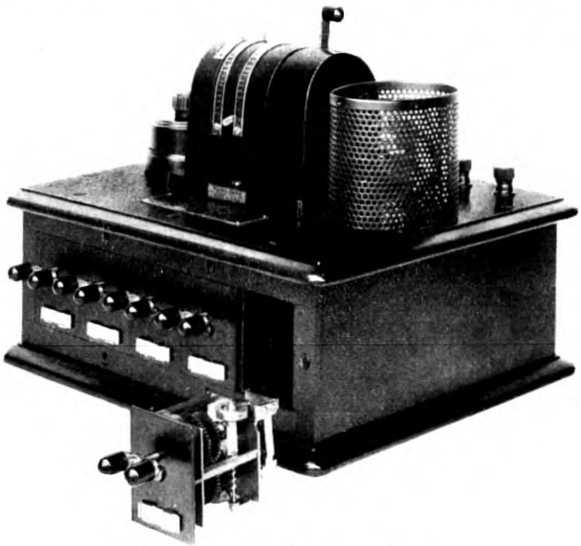


Fig. 1—Transmitter attachment

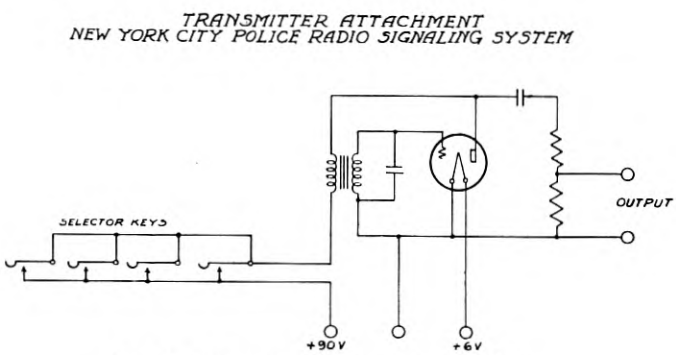


Fig. 2—Schematic of transmitter attachment

keys, which are connected in parallel, are in series with the plate winding of the oscillator coil. Operating any one of the calling keys opens and closes its contacts in a regular sequence determined by the code for which the key is set. Five group keys and a master key are provided. Each of the group keys is set to a certain code call which may be changed by resetting the key by means of a screw-driver. One of these group keys may be used as a call for the entire system so that by the operation of this key, every receiver is called simultaneously. The other keys may be used for the four main group calls. The master key is similar in appearance to a miniature cash register and by setting its levers to the proper code combination, any desired station may be called individually and this key may also be used for the sub-group combinations.

The output terminals of the oscillator of the transmitter attachment may be connected directly to the speech input equipment of a standard radio telephone transmitter in place of the microphone. The speech input amplifier is adjusted so that when the signaling attachment is used the maximum possible modulation is obtained. The sensitivity of the radio receiver is adjusted for reliable operation of the signaling system, which is sharply tuned to the frequency of the transmitter attachment. This frequency is 3,000 cycles and is so high that the volume of music or speech will be amply sufficient for easy reception but yet insufficient to operate the signaling system relays as only a relatively small proportion of the energy in normal speech or music occurs at frequencies in the vicinity of 3,000 cycles. Even should the relays be operated occasionally by excessive volume of speech or music the receiver signal lamp will not light unless the proper code call is sent.

Receiving Apparatus

Photographs of the receiving equipment are shown in Figs. 3, 4 and 5. At some of the outlying stations of the New York Police Department 110-volt DC power supply is available while at others 110 volt 60 cycle AC is provided. The radio receivers are made in two different types, one type for each kind of power supply. A schematic circuit of the DC type of receiver is shown in Fig. 6 and that of the AC type in Fig. 7. These two types are similar in all respects except such modifications as the different sources of power supply necessitate.

In the direct current type of receiver all of the vacuum tube filaments are connected in series, current being taken directly from the line through a filter to eliminate line noises due to generator hum and other causes. In parallel with each filament is connected a small switchboard lamp with a red cap mounted on the panel of the receiver cabinet. The resistance of each lamp filament is sufficiently greater than that of the vacuum tube filament so that the lamp will light only when the vacuum tube filament burns out or the tube is removed from its socket. In order to indicate when the power is turned on the receiver there is another lamp (green) mounted on the panel and connected across the 110 volt direct current line on the receiver side of the main switch so that when the switch is closed the lamp will light.

In the alternating current type of receiver the filaments of the

Fig. 3—Exterior view of receiving set. (The external appearance of the A.C. and D.C. models is the same)

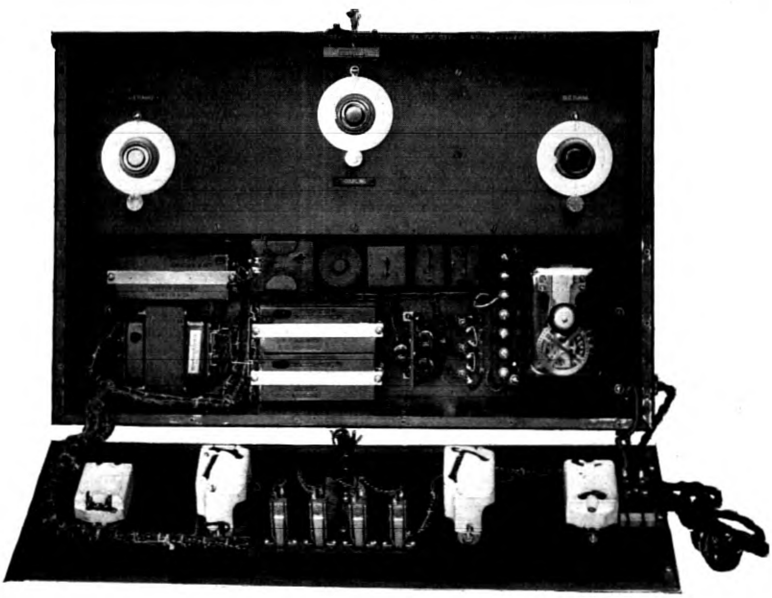


Fig. 4—View of A.C. model receiver showing tuning controls, and selector, and rectifying apparatus

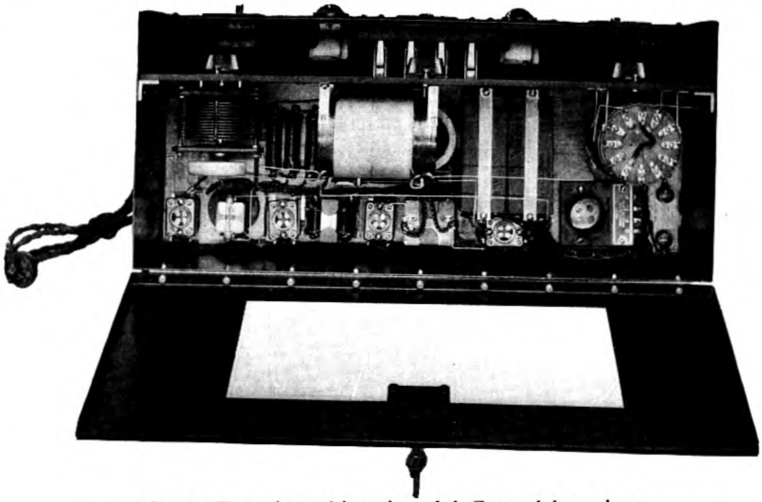


Fig. 5—Top view of interior of A.C. model receiver

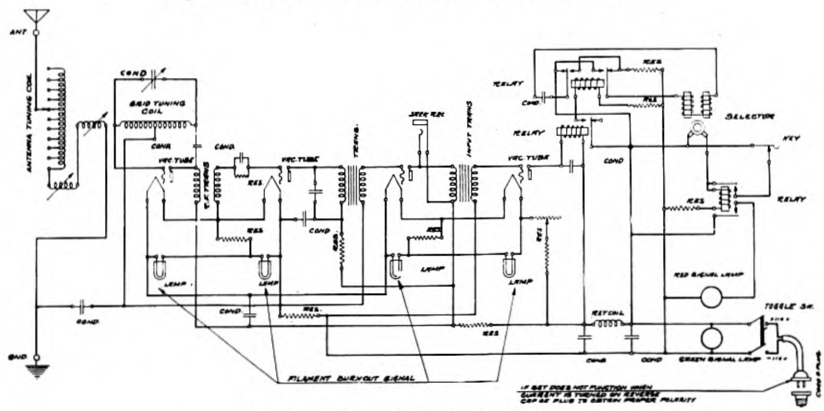


Fig. 6—Schematic of D.C. receiver

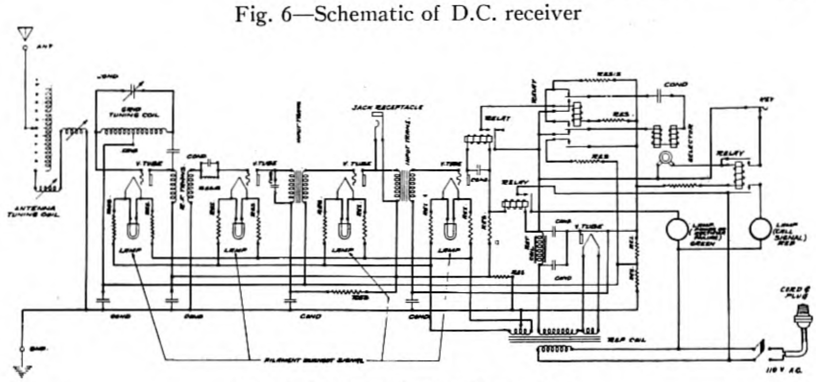


Fig. 7—Schematic of A.C. receiver

vacuum tubes are connected in parallel and are operated on alternating current. In order to take care of the filament warning lamps a resistance is provided in series with each tube filament so that when this filament burns out the voltage will rise sufficiently to light the corresponding lamp. To supply the necessary direct current for the plate potential of the vacuum tubes and for the operation of the relays, this receiver is provided with a power rectifier tube. The rectifier tube circuit is connected to the high voltage secondary winding of a power transformer and in connection with a filter system supplies to the receiver 110 volts direct current. As an indication of a burnt out rectifier tube filament there is provided a relay connected in series with the rectified current and a green signal lamp connected across the line in series with the contacts of this relay. The lamp will light when the power is turned on only if the rectifier tube is functioning properly. Thus it serves two useful purposes.

Both the DC and AC types of radio receivers are designed for operation from an open antenna, a double tuned, inductively coupled circuit being employed. The antenna circuit is tuned by means of taps on the loading coil and a small series inductance whose coupling to the loading coil is variable. An adjustable coupling coil is also included in the antenna circuit and serves as a sensitivity control. The secondary circuit to which this coil is coupled is tuned by means of a variable air condenser. All of these controls are on a panel inside the receiver and are inaccessible when the receiver cabinet is locked, thus insuring their remaining in adjustment.

Four "peanut" vacuum tubes are employed in the radio receiver. The first tube serves as a radio frequency amplifier. By means of a fixed radio frequency transformer sharply tuned to the frequency of the transmitting station, this tube is coupled to the second tube, which acts as a grid leak detector. This arrangement provides additional selectivity and more amplification than if a broadly tuned transformer were used, and is permissible since, in any given system, it is anticipated that the transmitter frequency will remain constant. To adapt the receiver to operate at any other transmitting frequency the transformers may be readily replaced by others of the proper characteristics. The third tube serves as an audio-frequency amplifier, being coupled to the detector tube by means of an audio-frequency transformer having a frequency characteristic suitable for the transmission of speech or music. The fourth tube serves as a rectifier and is coupled to the amplifier by means of a transformer, sharply tuned to the signaling frequency of 3,000 cycles. This frequency, as noted above, is sufficiently above the preponderant frequencies of normal speech or music to make the accidental operation of the signaling system a very remote possibility.

The normal plate current in the rectifier tube is a small fraction of a milliampere. Upon a signal being received, this current increases sufficiently to operate a relay connected in the plate circuit. The operation of this relay closes a circuit through the winding of another relay, which is the reversing relay for the operation of the selector controlling the red signal lamp which indicates a call.

The selector is the heart of the signaling system. This selector is shown in Fig. 8. It consists of a mechanism unit mounted upon a magnet unit, the whole being enclosed in a glass case for protection.

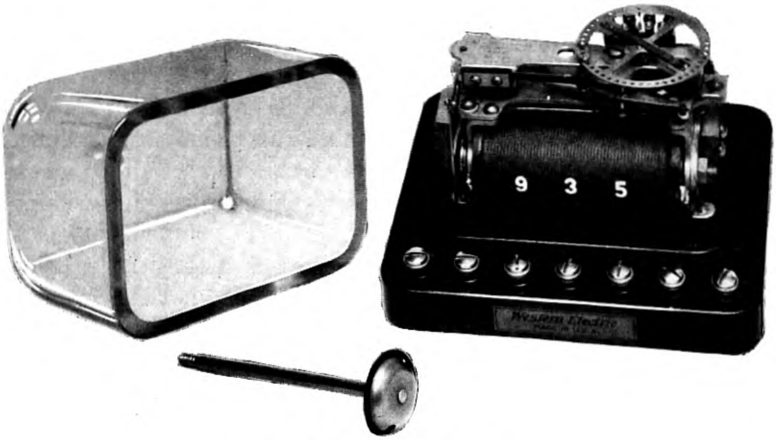


Fig. 8—Selector

The magnet windings are connected to a source of direct current of suitable voltage in series with a large fixed condenser and the contacts of the reversing relay mentioned above. When this relay is in either end position the circuit is completed through the condenser and the windings of the selector. As the power source is direct current it serves only to charge the condenser to its own potential. When the reversing relay is alternately operated and released, however, the repeated charging of this condenser in opposite directions sends pulses of current through the windings of the selector. This gives a rocking motion to the armature of the selector which motion is transmitted through a ratchet to the selector code wheel, advancing it against the tension of a spring one step for each movement of the armature. If the code wheel is kept from flying back during the pauses, two in number, which occur between groups of impulses, as is the case when the signal corresponds to the code setting of a particular selector, it will be advanced step by step until it reaches the point at which it makes signaling contact. For any other signal, however, it will be released at the pause between some two groups of pulses and will then immediately fly back to its initial position. To hold the code wheel during signal pauses a series of pins is arranged to engage a spring arm on the selector frame, their position thus determining the code of the particular selector. As the master key can be so operated as to send signals without any pause a general call can be made through all selectors simultaneously whatever their code setting. When the code wheel has rotated a distance corresponding to twenty-seven impulses, a spring contact mounted upon it makes contact with the first of a series of four stationary contacts A, B, C, and D. Two more impulses make contact with the second of the series, two more with the third, and two more, or a total of 33, with the fourth. Only one of these contacts is connected in any individual selector and four large groups are thus provided. All the transmitting keys are so arranged that about one second after the completion of their calling signal, they send a signal which restores all selectors under their control to normal.

By omitting one or both of the pauses between the three groups of impulses, it is apparent that each selector will respond to four and only four systems of pulses for each contact. On contact A, for example, the selector will close the signaling circuit if its individual call is sent, if the first pause only or the second pause only is omitted, or if both pauses are omitted and 27 consecutive impulses are sent. All of the selectors using any one of the four contacts may thus be grouped in several different ways. The total possible number of individual stations in each of the four large groups is somewhat over 200, or over 800 in the entire system, the exact number depending upon the grouping system which is employed. Each large group may be subdivided into a number of small groups having from 15 to 20 stations in each group, of which each station may belong to two groups if desired. In any case the number of consecutive impulses without pauses corresponding to the contact used will call all the stations in that large group, the sub-groups being formed by omitting one of the pauses. The system is thus capable of a high degree of flexibility.

The operation of the selector closes a circuit through the winding of a relay. This relay is of the slow operating type, this being necessary in order to avoid signals due to momentary contacts which are made by the selector with certain code combinations. The relay is provided with a holding contact so that after the selector has been restored to normal by the releasing impulse, it will remain operated until the person at the receiving station presses the key which opens the circuit of the holding contact. In the operated position, the relay completes the circuit of the red signal lamp on the panel of the receiver, thus indicating to the operator that he has been called. An audible signal may also be connected in parallel with it, using an additional relay if necessary to handle the heavy current required by a large gong.

In the exterior views of the receivers at the extreme left of the receiver is shown a jack into which the head telephones may be plugged. Plugging in these telephones does not interfere in any way with the operation of the signal system. Just to the right of the telephone jack is shown the red signal lamp for indicating when the receiver has been called. In the middle of the panel is a key which is used by the operator of the receiver to extinguish the red signal lamp when he takes up his head telephones. To the right of the key is the green signal lamp indicating when the power is on the receiver, and to the extreme right is the switch for turning the power on and off.

Wave Propagation in Overhead Wires with Ground Return

By JOHN R. CARSON

I

THE problem of wave propagation along a transmission system composed of an overhead wire parallel to the (plane) surface of the earth, in spite of its great technical importance, does not appear to have been satisfactorily solved.[1] While a complete solution of the actual problem is impossible, on account of the inequalities in the earth's surface and its lack of conductive homogeneity, the solution of the problem, where the actual earth is replaced by a plane homogeneous semi-infinite solid, is of considerable theoretical and practical interest. The solution of this problem is given in the present paper, together with formulas for calculating inductive disturbances in neighboring transmission systems.

The axis of the wire is taken parallel to the z-axis at height h above the xz-plane and passes through the y-axis at point O' as shown in Fig. 1 herewith. The "image" of the wire is designated by O''.

For $y>0$ (in the dielectric) the medium is supposed to have zero conductivity, while for $y<0$ (in the ground) the conductivity of the medium is designated by λ. The xz-plane represents the surface of separation between dielectric and ground.

We consider a wave propagated along the z-axis and the current, charge and field are supposed to contain the common factor $\exp(-\Gamma z+i\omega t)$, which, however, will be omitted for convenience in the formulas. The propagation constant, Γ, is to be determined. It is assumed, *ab initio*, as a very small quantity in c.g.s. units.[2]

In the ground ($y\leqq 0$) the axial electric force is formulated as the

[1] See Rudenberg, Zt. f. Angewandt, Math. u. Mechanik, Band 5, 1925. In that paper the current density in the ground is assumed to be distributed with radial symmetry. The resulting formulas are not in agreement with those of the present paper. Since this paper was set up in type I have learned that formulas equivalent to equations (26), (28), (31) for the *mutual impedance* of two parellel wires were obtained by my colleague, Dr. G. A. Campbell, in 1917. It is to be hoped that his solution will be published shortly.

[2] The simplifying assumptions introduced in this analysis are essentially the same as those employed and discussed in "Wave Propagation Over Parallel Wires: The Proximity Effect," *Phil. Mag.*, Vol. xli, April, 1921.

general solution, symmetrical with respect to x, of the wave equation; thus

$$E_z = -\int_0^\infty F(\mu) \cos x\mu \, e^{y\sqrt{\mu^2+i\alpha}} d\mu, \qquad y \leqq 0 \tag{1}$$

where

$\alpha = 4\pi\lambda\omega$,
$\lambda =$ conductivity of ground in elm. c.g.s. units,
$i = \sqrt{-1}$,
$\omega/2\pi =$ frequency in cycles per second.

(In the following analysis and formulas, elm. c.g.s. units are employed throughout).

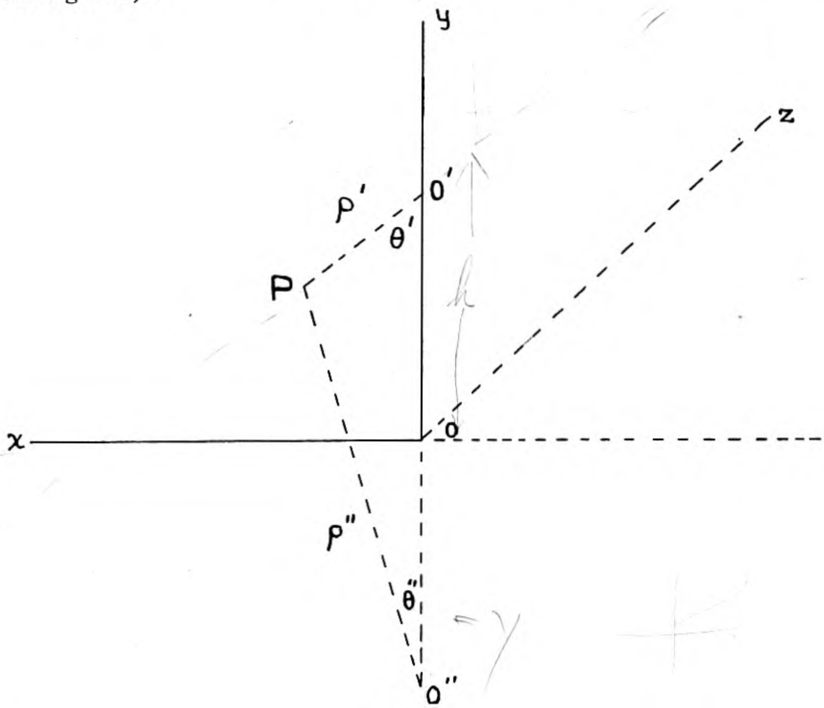


Fig. 1

Assuming that in the ground E_x and E_y are negligible compared with E_z, we have from the formula, curl $E = -\dfrac{\partial}{\partial t} H$,

$$i\omega H_x = -\frac{\partial}{\partial y} E_z$$

$$i\omega H_y = \frac{\partial}{\partial x} E_z.$$

Whence, in the ground

$$H_x=\frac{1}{i\omega}\int_0^\infty \sqrt{\mu^2+i\alpha}\,.F(\mu).\cos x\mu.e^{y\sqrt{\mu^2+i\alpha}}d\mu, \tag{2}$$

$$H_y=\frac{1}{i\omega}\int_0^\infty \mu.F(\mu).\sin x\mu.e^{y\sqrt{\mu^2+i\alpha}}d\mu, \tag{3}$$

it being understood that $y\leqq 0$. The function $F(\mu)$ in the preceding formulas is to be determined by the boundary conditions.

In the dielectric, H_x and H_y may be regarded as composed of two components; thus

$$H_x=H_x^0+H_x',$$
$$H_y=H_y^0+H_y',$$

where H_x^0, H_y^0 designate the field due to the current I in the wire, and H_x', H_y' the field of the ground current.

Neglecting axial displacement currents in the dielectric, and assuming that the wire is of sufficiently small radius so that the distribution of current over its cross section is symmetrical, we have

$$H_x^0=\frac{\cos\Theta'}{\rho'}.\,2I, \tag{4}$$

$$H_y^0=\frac{\sin\Theta'}{\rho'}.\,2I, \tag{5}$$

where $\rho'=\sqrt{x^2+(y-h)^2}$,

$$\cos\Theta'=\frac{h-y}{\rho'}, \tag{6}$$

$$\sin\Theta'=x/\rho'.$$

The secondary magnetic field H_x', H_y' is taken as

$$H_x'=\int_0^\infty \phi(\mu)\cos x\mu.e^{-y\mu}d\mu, \tag{7}$$

$$H_y'=-\int_0^\infty \phi(\mu)\sin x\mu.e^{-y\mu}d\mu. \tag{8}$$

At the surface of separation $y=0$, H_x^0, H_y^0 are expressible as the Fourier integrals

$$H_x^0=2I\int_0^\infty \cos x\mu.e^{-h\mu}d\mu, \tag{9}$$

$$H_y^0=2I\int_0^\infty \sin x\mu.e^{-h\mu}d\mu. \tag{10}$$

Also at the surface of separation of the two media ($y=0$), H_x and H_y must be continuous. Equating the values of H_x and H_y at $y=0$, as given by (2), (3) and by (7), (8) and (9), (10), we have

$$\frac{1}{i\omega}\sqrt{\mu^2+i\alpha}.F(\mu)=2I.e^{-h\mu}+\phi(\mu),$$

$$\frac{1}{i\omega}\mu.F(\mu)=2I.e^{-h\mu}-\phi(\mu),$$

whence

$$F(\mu)=\frac{i\omega e^{-h\mu}}{\sqrt{\mu^2+i\alpha}+\mu}4I, \tag{11}$$

$$\phi(\mu)=\frac{(\sqrt{\mu^2+i\alpha}-\mu)}{\sqrt{\mu^2+i\alpha}+\mu}e^{-h\mu}.2I, \tag{12}$$

which determines the functions $F(\mu)$ and $\phi(\mu)$.

Inserting the value of $F(\mu)$, as given by (11) in (1), the axial electric force E_z in the ground ($y \leqq 0$) and therefore the distribution of current density in the ground is expressed as a Fourier integral in terms of the frequency $\omega/2\pi$, the current I in the wire, the height h of the wire above ground, and the conductivity λ of the ground. Similarly the insertion of $\phi(\mu)$, as given by (12) in formulas (7) and (8) gives the magnetic field H_x, H_y in the dielectric. Thus

$$E_z(x,y)=E_z=-i4\omega I\int_0^\infty \frac{e^{-\mu h}}{\sqrt{\mu^2+i\alpha}+\mu}e^{y\sqrt{\mu^2+i\alpha}}\cos x\mu.d\mu, \qquad y\leqq 0. \tag{13}$$

This can be further simplified if we write

$$x'=x\sqrt{\alpha}$$

$$y'=y\sqrt{\alpha}$$

$$h'=h\sqrt{\alpha},$$

whence

$$E_z=-4\omega I\int_0^\infty (\sqrt{\mu^2+i}-\mu)e^{-h'\mu}.e^{y'\sqrt{\mu^2+i}}\cos x'\mu\, d\mu, \qquad y\leqq 0. \tag{14}$$

The axial electric force in the dielectric is now to be formulated. This is always derivable from a vector and a scalar potential; thus

$$E_z=-i\omega A_z-\frac{\partial}{\partial z}V, \tag{15}$$

where A_z is the vector potential of the axial currents, and V the scalar potential. Consequently,

$$E_z(x,y)-E_z(x,0)=-i\omega(A_z(x,y)-A_z(x,0))-\frac{\partial}{\partial z}(V(x,y)-V_0). \tag{16}$$

Here $E_z(x,0)$ is the axial electric intensity at the surface of the ground plane $(y=0)$, and

$$A_z(x,y)-A_z(x,0)=\int_0^y H_x(x,y)dy. \tag{17}$$

$V(x,y)-V_0$ is the difference in the scalar potential between the point x,y and the ground, which is due to the charges on the wire and on the surface of the ground. For convenience, it will be designated by V.

By means of (16) and the preceding formulas we get [3]

$$\begin{aligned} E_z=&-4\omega I\int_0^\infty(\sqrt{\mu^2+i}-\mu)e^{-(h'+y')\mu}\cos x'\mu\, d\mu \\ &-i\,2\omega I\log(\rho''/\rho')-\frac{\partial}{\partial z}V, \qquad y\geqq 0 \end{aligned} \tag{18}$$

where

$$\rho'=\sqrt{(h-y)^2+x^2}$$
$=$distance of point x,y from wire,
$$\rho''=\sqrt{(h+y)^2+x^2}$$
$=$distance of point x,y from image of wire.

The first two terms on the right hand side of (18) represent the electric force due to the varying magnetic field; the term $-\frac{\partial}{\partial z}V$ represents the axial electric intensity due to the charges on the surface of the wire and the ground. If Q be the charge per unit length, V is calculable by usual electrostatic methods on the assumption that the surface of the wire and the surface of the ground are equipotential surfaces, and their difference of potential is Q/C where C is the electrostatic capacity between wire and ground.[4]

II

By aid of the preceding analysis and formulas, we are now in a position to derive the propagation constant, Γ, and characteristic impedance, K, which characterize wave propagation along the system. Let z denote the "internal" or "intrinsic" impedance of the wire per

[3] As a check on this formula note that together with (14) it satisfies the condition of continuity of E_z at $y=0$.

[4] See "Wave Propagation Over Parallel Wires: The Proximity Effect," *Phil. Mag.*, Vol. xli, Apr., 1921.

unit length. (With small error this may usually be taken as the resistance per unit length of the wire.) The axial electric intensity at the surface of the wire is then zI. Equating this to the axial electric intensity at the surface of the wire as given by (18) and replacing $\partial/\partial z$ by $-\Gamma$, we have

$$zI = -4\omega I \int_0^\infty (\sqrt{\mu^2+i}-\mu)e^{-2h'\mu}d\mu$$
$$-i\,2\omega I.\log(\rho''/a)+\Gamma V. \tag{19}$$

Writing $V=Q/C$ and

$$i\omega Q=\Gamma I-GV=\Gamma I-\frac{G}{C}Q,$$

where G is the leakage conductance to ground per unit length, we have, solving for Γ,

$$\Gamma^2=(G+i\omega C)[z+i2\omega.\log(\rho''/a)]+4\omega\int_0^\infty (\sqrt{\mu^2+i}-\mu)e^{-2h'\mu}d\mu). \tag{20}$$

Writing this in the usual form

$$\Gamma^2=(R+iX)(G+i\omega C), \tag{21}$$

the characteristic impedance is given by

$$K^2=\frac{R+iX}{G+i\omega C} \tag{22}$$

and the *series impedance per unit length of the circuit* is

$$R+iX=Z=z+i2\omega.\log(\rho''/a)+4\omega\int_0^\infty (\sqrt{\mu^2+i}-\mu)e^{-2h'\mu}d\mu. \tag{23}$$

It will be observed that the first two terms on the right hand side of (23) represent the series impedance of the circuit *if the ground is a perfect conductor*; the infinite integral formulates the effect of the finite conductivity of the ground.

The *mutual impedance* [5] Z_{12} between two parallel ground return circuits with wires at heights h_1 and h_2 above ground and a separation x between their vertical planes is given by

$$Z_{12}=i2\omega.\log(\rho''/\rho')+4\omega\int_0^\infty (\sqrt{\mu^2+i}-\mu)e^{-(h_1'+h_2')\mu}\cos x'\mu\,d\mu, \tag{24}$$

[5] It will be noted that the mutual impedance is equal to the axial electric intensity at the axis of the second wire due to the varying magnetic field of unit current in the first wire and its accompanying distribution of ground current.

where

$$\rho''=\sqrt{(h_1+h_2)^2+x^2}$$
$$\rho'=\sqrt{(h_1-h_2)^2+x^2}$$
$$h_1'=h_1\sqrt{\alpha}$$
$$h_2'=h_2\sqrt{\alpha}$$
$$x'=x\sqrt{\alpha}.$$

From the preceding the series *self impedance* of the ground return circuit may be conveniently written as

$$Z=Z^0+Z' \tag{25}$$

and the *mutual impedance* as

$$Z_{12}=Z_{12}^0+Z_{12}' \tag{26}$$

where Z^0, Z_{12}^0 are the self and mutual impedances respectively, *on the assumption of a perfectly conducting ground*, and

$$Z'=4\omega\int_0^\infty(\sqrt{\mu^2+i}-\mu)e^{-2h'\mu}d\mu, \tag{27}$$

$$Z_{12}'=4\omega\int_0^\infty(\sqrt{\mu^2+i}-\mu)e^{-(h_1'+h_2')\mu}\cos x'\mu\, d\mu. \tag{28}$$

The calculation of the circuit constants and the electromagnetic field in the dielectric depends, therefore, on the evaluation of an infinite integral of the form

$$J(p,q)=J=\int_0^\infty(\sqrt{\mu^2+i}-\mu)e^{-p\mu}.\cos q\mu\, d\mu. \tag{29}$$

In terms of this integral

$$Z'=4\omega.J(2h'.0) \tag{30}$$

$$Z_{12}'=4\omega.J(h_1'+h_2',\, x'). \tag{31}$$

To the solution of the infinite integral $J(p,q)$ we now proceed.

III

The solution of equation (29), that is, the evaluation of $J(p,q)$ can be made to depend on the solution of the infinite integral

$$\int_0^\infty\sqrt{\mu^2+\alpha^2}.e^{-\beta\mu}d\mu$$

which has been worked out and communicated to me by R. M. Foster. It is

$$\frac{\alpha}{\beta}\left\{ K_1(\alpha\beta)+G(\alpha\beta) \right\}$$

where $K_1(x)$ is the Bessel function of the second kind and first order as defined by Jahnke und Emde, Funktionentafeln, pg. 93, and $G(x)$ is the absolutely convergent series

$$G(x)=\frac{x^2}{3}-\frac{x^4}{3^2.5}+\frac{x^6}{3^2.5^2.7}-\ \ldots .$$

On the basis of this solution, it is a straightforward though intricate and tedious process to derive the following solution for $J(p,q)$ of equation (29).

Writing $r=\sqrt{p^2+q^2}$ and $\Theta=\tan^{-1}(q/p)$, it is $J=P+iQ$

in which

$$\begin{aligned} P=&\frac{\pi}{8}(1-s_4)+\frac{1}{2}\left(\log\frac{2}{\gamma r}\right)s_2+\frac{1}{2}\Theta.s_2' \\ &-\frac{1}{\sqrt{2}}\sigma_1+\frac{1}{2}\sigma_2+\frac{1}{\sqrt{2}}\sigma_3, \end{aligned} \tag{32}$$

$$\begin{aligned} Q=&\frac{1}{4}+\frac{1}{2}\left(\log\frac{2}{\gamma r}\right)(1-s_4)-\frac{1}{2}\Theta.s_4' \\ &+\frac{1}{\sqrt{2}}\sigma_1-\frac{\pi}{8}s_2+\frac{1}{\sqrt{2}}\sigma_3-\frac{1}{2}\sigma_4. \end{aligned} \tag{33}$$

In these equations $\log\gamma$ is Euler's constant:

$\gamma=1.7811$, $\log\dfrac{2}{\gamma}=0.11593$, $\log\gamma=0.57722$ and σ_1, σ_2, σ_3, σ_4, s_2, s_2',

s_4, s_4', are infinite series defined as follows:

$$s_2=\frac{1}{1!2!}\left(\frac{r}{2}\right)^2\cos 2\Theta-\frac{1}{3!4!}\left(\frac{r}{2}\right)^6\cos 6\Theta+\ \ldots,$$

$$s_2'=\frac{1}{1!2!}\left(\frac{r}{2}\right)^2\sin 2\Theta-\frac{1}{3!4!}\left(\frac{r}{2}\right)^6\sin 6\Theta+\ \ldots,$$

$$s_4=\frac{1}{2!3!}\left(\frac{r}{2}\right)^4\cos 4\Theta-\frac{1}{4!5!}\left(\frac{r}{2}\right)^8\cos 8\Theta+\ \ldots$$

$$s_4' = \frac{1}{2!3!}\left(\frac{r}{2}\right)^4 \sin 4\Theta - \frac{1}{4!5!}\left(\frac{r}{2}\right)^8 \sin 8\Theta + \dots,$$

$$\sigma_1 = \frac{r\cos\Theta}{3} - \frac{r^5 \cos 5\Theta}{3^2.5^2.7} + \frac{r^9\cos 9\Theta}{3^2.5^2.7^2.9^2.11} - \dots,$$

$$\sigma_3 = \frac{r^3\cos 3\Theta}{3^2.5} - \frac{r^7\cos 7\Theta}{3^2.5^2.7^2.9} + \frac{r^{11}\cos 11\Theta}{3^2.5^2.7^2.9^2.11^2.13} - \dots,$$

$$\sigma_2 = \left(1+\frac{1}{2}-\frac{1}{4}\right)\frac{1}{1!2!}\left(\frac{r}{2}\right)^2 \cos 2\Theta$$

$$-\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}-\frac{1}{8}\right)\frac{1}{3!4!}\left(\frac{r}{2}\right)^6 \cos 6\Theta + \dots,$$

$$= \frac{5}{4}s_2 \text{ approximately,}$$

$$\sigma_4 = \left(1+\frac{1}{2}+\frac{1}{3}-\frac{1}{6}\right)\frac{1}{2!3!}\left(\frac{r}{2}\right)^4 \cos 4\Theta$$

$$-\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}-\frac{1}{10}\right)\frac{1}{4!5!}\left(\frac{r}{2}\right)^8 \cos 8\Theta + \dots.$$

$$= \frac{5}{3}s_4 \text{ approximately.}$$

It is to be regretted that the foregoing formulas appear so complicated. The series, however, are very rapidly convergent and for $r \leqq 2$ only the two leading terms of each series need be retained. For $r \leqq 1$, only the leading terms are of importance.

For the important range $r \leqq 1/4$,

$$P = \frac{\pi}{8} - \frac{1}{3\sqrt{2}} r\cos\Theta + \frac{r^2}{16}\cos 2\Theta\left(0.6728 + \log\frac{2}{r}\right) + \frac{r^2}{16}\Theta \sin 2\Theta, \quad (34)$$

$$Q = -0.0386 + \frac{1}{2}\log\left(\frac{2}{r}\right) + \frac{1}{3\sqrt{2}} r\cos\Theta. \quad (35)$$

For $r > 5$ the following asymptotic expansions, derivable from (29) by repeated partial integrations, are to be employed.

$$P = \frac{1}{\sqrt{2}}\frac{\cos\Theta}{r} - \frac{\cos 2\Theta}{r^2} + \frac{1}{\sqrt{2}}\frac{\cos 3\Theta}{r^3} + \frac{3}{\sqrt{2}}\frac{\cos 5\Theta}{r^5} \dots, \quad (36)$$

$$r > 5.$$

$$Q = \frac{1}{\sqrt{2}}\frac{\cos\Theta}{r} - \frac{1}{\sqrt{2}}\frac{\cos 3\Theta}{r^3} + \frac{3}{\sqrt{2}}\frac{\cos 5\Theta}{r^5} - \dots, \quad (37)$$

For large values of $r (r > 10)$, these reduce to

$$J = \frac{1+i}{\sqrt{2}} \frac{\cos \Theta}{r} - \frac{\cos 2\Theta}{r^2}, \; r > 10. \tag{38}$$

In view of the somewhat complicated character of the function in the range $1/4 \leqq r \leqq 5$ the curves shown below have been computed. These show $J = P + iQ$ as a function of r for $\Theta = 0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}$. By interpolation it is possible to estimate with fair accuracy the value of the functions for intermediate values of Θ.

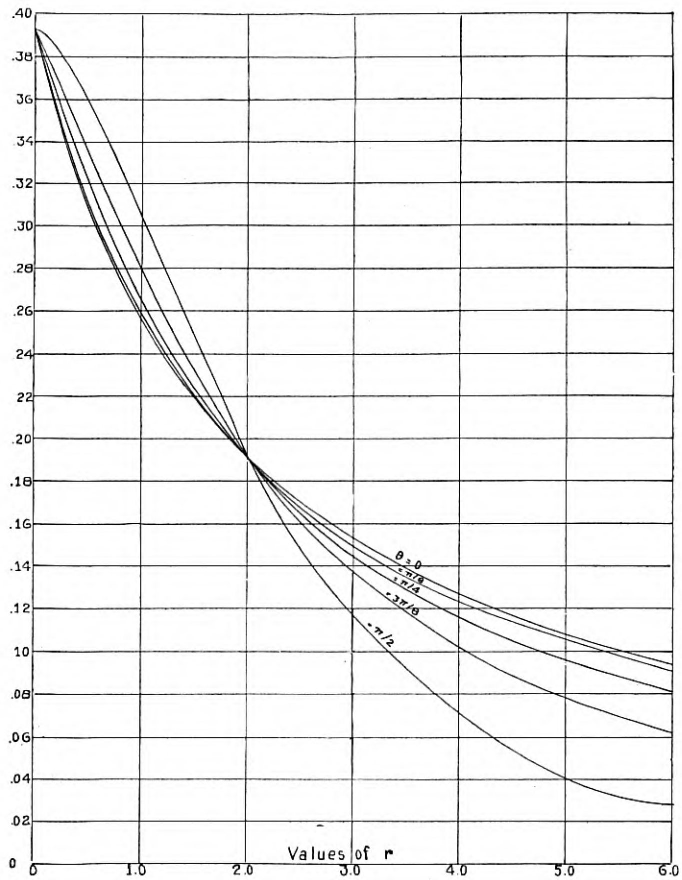


Fig. 2 P = real part of J

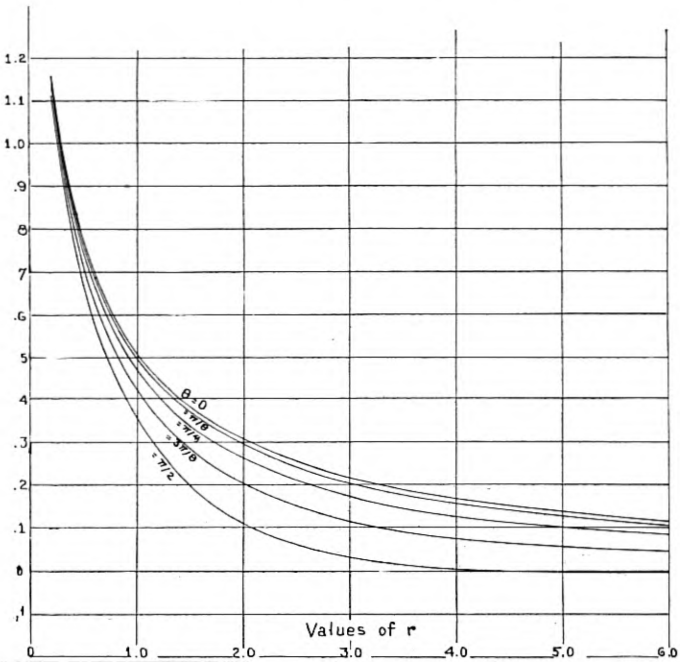


Fig. 3 Q = imaginary part of J

IV

The theory and formulas of the preceding sections will now be reviewed and summarized as regards their principal applications to technical transmission problems where the ground forms, in whole or part, the "return" part of the circuit. In such problems we are interested in the electric intensity in the dielectric and in the ground, and in particular in the self impedance and mutual impedances of ground return circuits. The calculation of these quantities is provided for by the general analysis and the solution of the infinite integral J. Reference should be made to Fig. 1 shown in section I for the geometry of the system and coordinate system employed.

1. *The Axial Electric Intensity E_z in the Dielectric.* (See equations (15) and (18)).

$$E_z = -\frac{\partial}{\partial z}V - (i2\omega \log (\rho''/\rho') + 4\omega J)I$$

where

$$\rho' = \sqrt{(h-y)^2+x^2}$$

= distance of point in dielectric *from axis* of wire.

$$\rho'' = \sqrt{(h+y)^2+x^2}$$

= distance of point in dielectric *from image* of wire.

$$r = \rho''\sqrt{\alpha}$$

$$\Theta = \sin^{-1}(x/\rho'')$$

$$\alpha = 4\pi\lambda\omega.$$

These values of r and Θ are, of course, to be employed in calculating $J = P + iQ$ from the formulas and curves of the preceding section. As a special case *the electric intensity at the surface of the earth is*

$$E_z = -4\omega JI$$

$$\rho'' = \sqrt{h^2+x^2}$$

$$r = \rho''\sqrt{\alpha}$$

$$\Theta = \sin^{-1}(x/\rho'')$$

2. *Self Impedance of Ground Return Circuit.* (See equations (25), (27), (30)).

$$Z = Z^0 + 4\omega J$$

Z^0 = self impedance with perfectly conducting ground.

$$r = 2h\sqrt{\alpha}$$

$$\Theta = 0.$$

3. *Mutual Impedance of Ground Return Circuits.* (See equations (26), (28), (31)).

$$Z_{12} = Z^0_{12} + 4\omega J$$

Z^0_{12} = mutual impedance with perfectly conducting ground.

$$r = \sqrt{\alpha}\sqrt{(h_1+h_2)^2+x^2} = \rho''\sqrt{\alpha}$$

$$\Theta = \sin^{-1}(x/\rho'').$$

The axial electric intensity E_z in the ground ($y<0$) is given by equation (1), and the subsequent analysis, whence

$$E_z = -4\omega I \int_0^\infty (\sqrt{\mu^2+i}-\mu).\cos x'\mu.e^{-(h'\mu+g'\sqrt{\mu^2+i})}d\mu$$

where, as before

$$x' = x\sqrt{\alpha}$$

$$h' = h\sqrt{\alpha}$$

and

$$g' = \sqrt{\alpha} \text{ times the depth below the surface of the ground.}$$
$$= g\sqrt{\alpha}.$$

The integral can undoubtedly be evaluated in somewhat the same way as (29) and can in any case be numerically computed without much difficulty. Owing, however, to the secondary technical interest in the electric intensity below the surface of the earth, the detailed solution has not been undertaken, nor has the magnetic field been worked out.

V

The practical utility of the preceding theory and formulas will now be illustrated by a brief sketch of their application to two important transmission problems.

The Wave Antenna

When a transmission line with "ground return" is employed as a radio receiving antenna it is called a wave antenna. The theory and design of such an antenna requires a knowledge of the transmission characteristics of the ground return circuit, which are calculable, as shown above, from the geometry and constants of the overhead wire, together with $Z' = 4\omega J$, which may be termed the "ground return" impedance.

We assume that the wire is approximately 30 ft. above the ground ($h = 10^3$) and that the frequency is 5.10^4 c.p.s. corresponding to the frequency employed in Trans-Atlantic radio communication. The ground conductivity λ is exceedingly variable, depending on the locality and weather conditions. Calculations of Z' will therefore be made for two extreme cases, $\lambda = 10^{-12}$ and $\lambda = 10^{-14}$ which should cover the range of variation encountered in practice.

For $\lambda = 10^{-12}$,

$$\sqrt{\alpha} = \sqrt{4\pi\lambda\omega} = 2.10^{-3}$$

and for $\lambda = 10^{-14}$,

$$\alpha = 2.10^{-4}.$$

Correspondingly, $r=2h\sqrt{\alpha}$ has the values 4.0 and 0.4, respectively. Reference to the preceding formulas and curves for J, for $r=4.0$ and $r=0.4$, give

$$J=0.126+i\ 0.168, \qquad \lambda=10^{-12}$$

$$J=0.323+i\ 0.871, \qquad \lambda=10^{-14}$$

whence the corresponding values of Z' are

$$Z'=4\omega.\ (0.126+i\ 0.168),$$

$$Z'=4\omega.\ (0.323+i\ 0.871).$$

These are the "ground return" impedances per unit length in elm. c.g.s. units; to convert to *ohms per mile* they are to be multiplied by the factor 1.61×10^{-4}. Consequently setting $\omega=\pi.10^5$, we get

$$Z'=6.44\pi(1.3+i\ 1.7), \qquad \lambda=10^{-12}$$

$$Z'=6.44\pi\ (3.2+i\ 8.7), \qquad \lambda=10^{-14}.$$

Comparison of these formulas shows that an hundred-fold increase in the resistivity of the ground increases the resistance component of the ground return impedance by the factor 2.5 and increases its reactance only five-fold. That is to say, the ground return impedance is not sensitive to wide variations in the resistivity of the earth, a fortunate circumstance in view of its wide variability and our lack of precise information regarding it.

Induction from Electric Railway Systems

A particularly important application of the preceding analysis is to the problems connected with the disturbances induced in parallel communication lines by alternating current electric railways. Assuming the frequency as 25 c.p.s., we have corresponding to $\lambda=10^{-12}$ and $\lambda=10^{-14}$,

$$\sqrt{\alpha}=0.45\times10^{-4} \text{ and } 0.45\times10^{-5}.$$

Taking the height of the trolley wire as approximately 30 ft., $h=10^3$ and assuming the parallel telephone as the same height above ground and separated by approximately 120 ft., $x=4.10^3$, and

$$r=\sqrt{\alpha}\sqrt{(2h)^2-x^2}$$

$$=4.47\times10^3\sqrt{\alpha},$$

and corresponding to the values of α taken above

$$r=0.2 \text{ and } 0.02 \text{ in round numbers,}$$

while

$$\Theta=\sin^{-1}\frac{4}{\sqrt{20}}=63^\circ\,30' \text{ approximately.}$$

For both cases, therefore, we can employ, in calculating $J=P+iQ$, the approximate formulas,

$$P=\frac{\pi}{8}-\frac{1}{3\sqrt{2}}r.\cos\Theta+\frac{r^2}{16}\cos 2\Theta\left(.6728+\log\frac{2}{r}\right)+\frac{r^2}{16}\Theta\sin 2\Theta,$$

$$Q=-0.0386+\frac{1}{2}\log\left(\frac{2}{r}\right)+\frac{1}{3\sqrt{2}}r\cos\Theta.$$

For $\lambda=10^{-12}$ and $r=0.2$, this gives

$$J=0.369+i\ 1.135$$

and

$$Z'_{12}=4\omega(0.369+i\ 1.135).$$

The foregoing assumes that the only return conductor is the ground. If, however, an equal and opposite current flows in the rail we must subtract from the foregoing mutual impedance, the mutual impedance between rail and telephone line; that is, the mutual impedance Z'_2 between the telephone line and a conductor at the surface of the earth. For this case

$$\rho''=\sqrt{h^2+x^2}=4.12\times 10^3$$

$$\Theta=\sin^{-1}\frac{4}{\sqrt{17}}=76^\circ$$

$$\cos\Theta=0.242,\ r=0.184 \text{ for } \lambda=10^{-12}.$$

The corresponding value of J is

$$J=0.378+i\ 1.165$$

and the *resultant* mutual impedance between railway and parallel telephone line is,

$$Z'_{12}=4\omega(0.369-0.378+i\ (1.135-1.165))$$
$$=4\omega(0.009-i\ 0.030).$$

The very large reduction in mutual impedance, due to the current in the rail, is striking.

For the case of $\lambda=10^{-14}$, the corresponding calculations give

$$Z'_{12}=4\omega(0.391+i\ 2.27)$$

with *no current in rail*, and

$$Z'_{12}=4\omega(-0.001-i\ 0.002)$$

with *equal and opposite current in rail.* It is evident from these figures that the reduction in mutual impedance, due to the current in the rail, is practically independent of the ground conductivity, at least at the separation specified.

Electrode Effects in the Measurement of Power Factor and Dielectric Constant of Sheet Insulating Materials

By E. T. HOCH

SYNOPSIS: It is shown that, aside from the guard ring type of electrode which can only be used with certain special types of measuring circuit, the most accurate results can probably be obtained by the use of equal foil electrodes and making proper allowance for the edge effects. From the standpoint of convenience, mercury electrodes and foil electrodes of unequal size have certain advantages. It is believed that the results obtained with these two types are also sufficiently accurate for most purposes when the corresponding corrections are applied.

INTRODUCTION

WHEN it is desired to determine the dielectric constant and power factor of an insulating material, the first problem that presents itself is that of finding a suitable means of applying a potential to the material in question. In order to accomplish this, a sample of the material must, in general, be placed between two conductors or electrodes to which the desired potential is applied. The size, shape and manner of application of these electrodes affect directly the quantities to be measured from which the dielectric constant and power factor are derived. Therefore, unless proper allowance can be made for them, these features of the electrodes will affect the values obtained for the properties of the insulation.

It is the purpose of this paper to discuss only the part played by the electrodes in the measurement of power factor and dielectric constant and not to discuss complete methods for measuring these quantities. Experimental data will be presented to show the magnitude of the various effects discussed.

If any form of test is to be of general use for determining the properties of any material, there are certain fundamental requirements which must be fulfilled. First, the method should lead to exact reproducibility of results. That is, a test on a given sample of material should lead to the same result regardless of when, where or by whom the test is made. Second, the result obtained should be the correct result, that is, the absolute accuracy should be high. Third, the method should be as convenient as possible to use. If it is not, the method loses its practical value to a large extent since the tendency will be for most people to use a more convenient method even at the expense of accuracy and reproducibility.

Sources of Error

There are certain sources of error inherent in all electrodes which affect both the reproducibility and the accuracy of the results obtained. First, we have the question of contact between the electrode and the sample. If the electrode does not make perfect contact over its entire area with the sample, the result obtained is not a true value for the material of the sample, but is a resultant, depending upon the amount and nature of the material filling the gap. Air spaces between the electrode and the sample have a marked effect on the apparent dielectric constant. An air-gap .001 in. thick in series with a sample having a dielectric constant of 5 will have the same effect on the capacitance as increasing the thickness of the sample by .005 in. If the actual thickness of the sample is .05 in. this results in an error of nearly 10% in the value of dielectric constant. The power factor will also be reduced and the loss factor or product of power factor and dielectric constant will be reduced by the factor $\left(\frac{50}{55}\right)^2$, or about 17%. Thus it is evident that the elimination of all gaps between the electrode and the sample is one of the first requirements both for reproducibility and accuracy.

A second effect inherent in all electrodes which is a source of error unless properly allowed for, is the so-called fringing of the electrostatic flux, that is, the lines of force tend to spread out and include an area of the sample greater than that of the electrode. This is

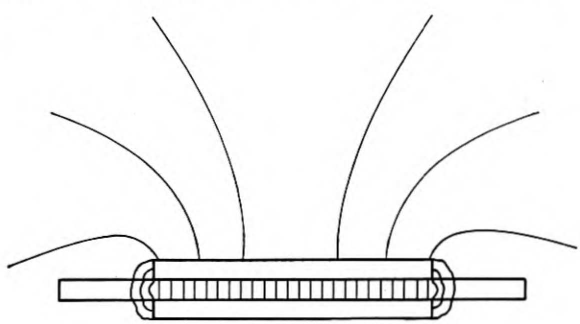


Fig. 1

illustrated in Fig. 1. So far as the flux which is confined entirely to the material of the sample is concerned, this produces an error in the dielectric constant, which involves a determination of the effective area of the sample, but not in the power factor which is independent of the area. However, there are some lines of force which terminate on the vertical surfaces of the electrode and whose paths are partly through the sample and partly through air. These introduce a slight error into the power factor also. They also make the capacitance depend to a slight extent on the thickness of the electrodes. These edge effects vary with the thickness of the sample and also with the ratio of the perimeter to the area of the electrode and hence with its size and shape. They are also increased materially when one electrode is larger than the other.

The third inherent source of error is the capacitance from the ungrounded electrode[1] to earth due to lines of force which pass out in all directions other than through the sample. This also is illustrated in Fig. 1. This increases the measured capacitance by an amount depending somewhat on the nature and position of surrounding objects. If this capacitance is due to flux passing entirely through air it makes no difference in the dielectric loss,[2] but if the path of the flux includes other material such as the wood, brick, plaster, etc., in the walls and floor of the room as is often the case, it may add an appreciable loss.

We will now consider the above errors and also the question of convenience as applied to certain specific types of electrodes. The following types will be considered.

1. Plain metal electrodes.
2. Mercury electrodes.
   a—Confining ring of metal.
   b—Confining ring of insulating material.
3. Foil electrodes.
   a—Both same size as sample.
   b—Both same size, but smaller than the sample
   c—One materially larger than the other.
4. Conducting paint electrodes.
5. Fixed gap electrodes.

Plain Metal Electrodes

One of the simplest forms of electrode would be two similar blocks of metal between which the sample would be placed. If the surfaces of both the electrode and sample were true planes, this would be fairly satisfactory. However, as the surfaces of samples of insulating material usually available for test are not true planes, the air-gap

[1] Assuming one electrode to be grounded as is usually the case.

[2] The apparent power factor is reduced by the increase in capacitance but the loss factor is not affected since the dielectric constant is increased to the same extent that the power factor is reduced.

error is usually prohibitive and makes this type of electrode practically useless.

Sometimes electrodes of this type are amalgamated and flooded with mercury and then pressed onto the sample, the excess mercury being brushed away.

This is an improvement but still leaves considerable uncertainty as to the degree of contact obtained.

Mercury Electrodes

Primarily on account of the ease with which a liquid will conform to the contour of an irregular surface, mercury has frequently been used as an electrode material. The usual procedure is to float the sample in a tray of mercury which serves as the lower electrode. A confining ring of some form is then placed on top of the sample into which a pool of mercury is poured which serves as the upper electrode.

When transparent samples are floated in this way it is observed that if a sample is simply laid flat upon the surface of the mercury, con-

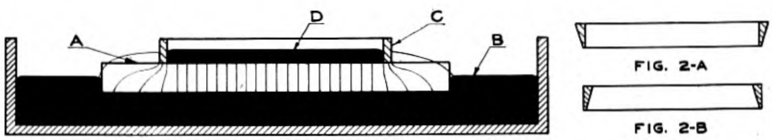


Fig. 2

siderable air is trapped between the sample and the mercury. However, if the sample is lowered obliquely on to the surface of the mercury the air can usually be eliminated. When the mercury has been poured on top of the sample it is impossible to see whether the air has been eliminated from the upper surface of the sample or not. This is sometimes considered a serious objection to the use of this form of electrode especially when used with very thin samples. It may also be questioned whether or not the mercury completely fills the angle between the sample and the inside surface of the ring.

Another point to be considered in the use of this type of electrode is whether the confining ring should be of conducting or of insulating material. The ideal condition, of course, would be to have the confining ring eliminated or, what is the same thing, to have it infinitely thin. Practically, however, something is required to confine the mercury to a definite area. Consider for a moment the arrangement shown in Fig. 2. *A* is the sample under test, *B* is the mercury constituting the lower electrode, C is the confining ring and D the mercury of the upper electrode. Consider first the ideal case where the ring is of finite height but infinitely thin. This corresponds to the condition illustrated in Fig. 1 except that the fringing is increased due to the size of the lower electrode. Suppose now that the ring is made of insulating material and of appreciable thickness. Some of the lines of force then pass through the material of the ring and the results become dependent on the nature and amount of this material. On the other hand, suppose the ring is made of metal and of appreciable thickness. This, of course, increases the effective area of the electrode but this is easily remedied by making the outside diameter of the ring equal to the size of electrode desired. However, that part of the surface which is actually covered by the ring is subject to the same error as in the case of plain metal electrodes, namely, that if the surface of the sample is not a true plane there will be a gap between the ring and the sample. One remedy for this is to make the ring thin enough so that the area affected is negligible compared with the total. This may not seem to be consistent with suitable rigidity. However, if a ring of at least 4½ inches diameter is used it can be as much as 1/16 inch thick and therefore quite rigid without serious error. In this case the area of the ring is about 5.6% of the total area and assuming a 10% error due to the air-gap as in the case of the plain electrodes the result would be a net error of .56% in the dielectric constant and a correspondingly slight error in the power factor.

Another means of reducing this error without sacrificing the rigidity of the ring is to bevel the edge of the ring either on the outside or inside as shown in Figs. 2-A and 2-B, respectively. In this way the area covered by the ring can be reduced to less than 1% of the total area and the air-gap error to less than 0.1%. However, if the ring is beveled on the outside the outer surface of the ring is no longer perpendicular to the surface of the sample and a slight increase in the stray field from this surface is produced. On the other hand, when the ring is beveled on the inside, the angle between the inner surface of the ring and the sample is less than a right angle and any tendency of the mercury not to fill this angle will be increased. However, both of these factors are probably but little affected by the slight bevel which is sufficient to reduce the thickness of the edge to a small value.

Another form of mercury electrode designed especially to eliminate all errors due to air bubbles has been used by Dye and Hartshorn for measuring the dielectric constant of mica in very thin sheets.[3]

[3] Proceedings of the Physical Society of London, December 15, 1924.

In this arrangement the sample is clamped in a vertical position between two ebonite plates. These plates are recessed in such a way that a closed cell is formed on each side of the sample. An air vent is provided at the top of the cell and the mercury is introduced through a capillary U tube connected to the bottom of each cell. Thus the mercury is caused to rise slowly along the surface of the sample displacing the air completely. Also a slight head of mercury can be maintained in order to force it into all angles. However, this electrode is subject to the errors of the confining ring of insulating material in an exaggerated form on account of the mercury electrodes being entirely enclosed by the ebonite clamping plates. This would unquestionably lead to appreciable errors in the power factor especially in the case of thick samples of low loss material.

Foil Electrodes

Another form of electrode which has been widely used consists simply of a sheet of tinfoil applied to either side of the sample usually with a thin film of wax or petrolatum to serve as an adhesive. This has the advantage that the thickness of the electrode can be made negligible thereby practically eliminating the error due to the field from the vertical surface of the electrode passing partly through air and partly through the sample. The capacitance from the upper surface of the electrode to ground however is not eliminated.

The question of the size of the electrodes also arises. If both the electrodes are extended entirely to the edge of the sample the edge correction for the capacitance is greatly reduced since the fringing all takes place in air having a dielectric constant of 1 instead of the higher dielectric constant of the sample. On the other hand the fringing through the air does produce a small effect on the power factor which does not exist when the fringing is all through the sample. However, the biggest objection to this arrangement from the practical standpoint probably is that the samples very frequently are not uniform in thickness near the edge and this makes it difficult to determine the effective thickness of the sample.

Another possibility is to make both electrodes the same size but smaller than the sample. This should result in a comparatively small edge correction but requires careful manipulation to insure the electrodes being exactly opposite each other. The simplest arrangement from a convenience standpoint is to have one large and one smaller electrode. This however, results in a further increase in the edge correction.

CONDUCTING PAINT ELECTRODES

Another form of electrode which might be considered as a variation of the foil electrode consists of a coating of conducting paint on either side of the sample. In general the conditions are the same as with foil electrodes. A possible advantage might be more intimate contact with the sample and a possible disadvantage is that the film may have sufficient resistance to materially affect the power factor measurement. Many metallic paints are almost entirely non-conducting and are therefore entirely unsuitable for this purpose. If suitable conductivity is obtained, the discussion of foil electrodes given above may be applied to this type also.

FIXED GAP ELECTRODES

Another type of electrode occasionally referred to [4] consists essentially of a parallel plate air condenser the capacity of which is measured first alone, and then with the sample of insulating material inserted in the air-gap but not necessarily filling the gap completely. From these measurements and the known dimensions of the air condenser and sample the dielectric constant and power factor of the sample can be computed. This arrangement is capable of high accuracy if the dimensions of the sample and the thickness of the gap are known with sufficient accuracy. The computations are not as simple as for the other electrodes referred to and a slight error in determining the thickness of the sample or gap results in a much larger error in the final results. Therefore, this method does not seem to offer any marked advantages over the simpler forms.

EVALUATION OF ERRORS

Having now discussed in a general way the various types of errors and their probable effect in connection with various types of electrodes, an attempt will be made to determine by experiment the magnitude of the more important of these errors with respect to certain definite types of electrodes. From the discussion already given, it appears that some form of the mercury or foil electrode should be the most suitable for general use. Hence this investigation will be confined to these two general types.

The first question, namely, that of reproducibility, is one which cannot be determined by a single observer except as it applies to his own particular method of manipulation. Since the results obtained

[4] A. Campbell, Proceedings of the Royal Society, Volume 78, Page 196 and Dye and Hartshorn Local Citation.

may depend considerably on the skill and patience exercised in the handling of the samples and electrodes, they may vary considerably with different observers. Hence, a comprehensive discussion of this point is beyond the scope of this paper. It has been the experience of the writer, however, that there is little choice between the two in this respect and that a decision between them rests primarily on other factors.

Since the magnitude of the ground capacitances and fringing effects both for foil and for mercury confined by a metal ring can be determined from a single series of tests, such an experiment will now be described.

Samples of insulating material 6 inches square were entirely coated on both sides with tinfoil using petrolatum as an adhesive. After the foils were in place, a 4½ inch circle was described on each foil from the center of the square and cut through so that the inner and outer portions were not in electrical contact, although the separation between them was very small. This left two 4½ inch disc electrodes L and M, on opposite sides of the sample surrounded by the annular pieces N and O respectively. When the inner and outer sections are connected, we have the condition of foil electrodes covering the entire surface of the sample. Then if N is used as a guard ring, it is possible to obtain measurements between L and M under uniform field conditions with all fringing effect eliminated. If N is removed and M and O are connected, we have the condition of one large and one smaller electrode. If a metal ring 4½ inches outside diameter is placed on the foil, we have a condition similar to that of a mercury electrode confined by a metal ring. If both N and O are removed, we have the case of two equal foil electrodes smaller than the sample. All of these variations can be obtained without any variation whatever in the conditions of contact between L and M and the sample and therefore are directly comparable.

The method of making these measurements [5] by means of a completely shielded capacitance and conductance bridge [6] is the same as that described by Campbell for the measurement of direct capacitance and will not be described here. The measurements were made at a frequency of 1,000 cycles as fewer difficulties are encountered than at radio frequencies and the general results are the same for any frequency. The capacitances were balanced to 0.1 mmf. or better, and the conductances to 0.0001 micro-mho.

[5] G. A. Campbell, *Bell System Technical Journal*, July, 1922 and *Journal of the Optical Society of America and Review of Scientific Instruments*, August, 1922.

[6] G. A. Campbell, *Electrical World*, 43, 1904, 647–649.

The complete series of measurements made on the samples discussed above was as follows:

1. Grounded capacitance of L+N to M+O. This includes the stray capacitance of the upper electrode to ground and also any fringing effect in the air around the edges of the sample.
2. Direct capacitance of L+N to M+O. This eliminates the stray capacitance to ground but includes the above fringing, therefore, 2 minus 1 gives the stray capacitance of the 6″ square electrode to ground.
3. Direct capacitance of L+N to ground using M+O as a shield. This should check 2 minus 1 above.
4. Direct capacitance from L to M using N as a guard ring and eliminating all fringing and ground capacitance.
5. Direct capacitance from L to M+O with N removed. This includes the fringe effect from a small to a large electrode but eliminates the capacitance of L to ground. Hence, 5 minus 4 gives the fringe effect.
6. Grounded capacitance of L to M+O with N removed. This includes both fringe effect and capacitance of L to ground. Hence, 6 minus 5 gives the capacitance of L to ground.
7. Same as 6 with 4½″ metal ring on top of foil L. 7 minus 6 gives the added capacitance due to the ring.
8. Direct capacitance of L to ground using M+O as shield. This should check 6 minus 5 above.
9. Direct capacitance of L to M with N and O removed. This includes the fringe effect between equal electrodes but eliminates the capacitance of L to ground.
10. Grounded capacitance of L to M with N and O removed. This includes the fringe effect and ground capacitance for equal electrodes. 10 minus 9 equals capacitance of L to ground with equal electrodes.
11. Direct capacitance of L to ground using M as shield. This should check 10 minus 9 above.

The results of these measurements on a number of samples including several thicknesses of phenol fibre, hard rubber and glass are given in Table I. In all cases, the capacitance of the leads was measured separately and deducted. The differences between readings as indicated above, and the corresponding check readings are tabulated in Table II.

In using the results tabulated in Table II the directly measured values of stray capacitance to ground (items 2, 6, 12) should be considered much more reliable than those determined by differences (items 1, 5, 11). The degree of agreement between the two should be considered more as a check on the accuracy of the individual values from which the latter are derived than as a check on the former.

Hence, we may consider 3.3 mmf. (item 2, Table II) as the stray capacitance for the 6 inch square and about 2.3 mmf. (item 12) for the 4½ inch circle. This checks fairly well with the theoretical value of $\frac{R}{\pi}$ (CGS units). If the square is considered equivalent to a circle of equal area the value of $\frac{R}{\pi}$ is equivalent to about 3.03 mmf. For the 4½ inch circle the value of $\frac{R}{\pi}$ is 2.0 mmf. The measured values should be somewhat higher than the theoretical since the shielded bridge and other apparatus comprise a considerable mass of grounded metal at no great distance from the sample.

The fringe effect for the equal circular electrodes may be compared with values computed from Kirchhoff's formula

$$C=\frac{r}{4\pi}\left(\log_e\frac{16\pi(b+t)r}{b^2}+\frac{t}{b}\log_e\frac{b+t}{t}-3\right),$$

where C is the fringe effect and r and t are the radius and thickness respectively of the electrodes and b is their separation, all in CGS units. For very thin electrodes this reduces to

$$C=\frac{r}{4\pi}\left(\log_e\frac{16\pi r}{b}-3\right).$$

Values for this expression reduced to a percentage correction are listed as item 13 in Table II and are plotted in Fig. 3. It will be noted that the observed effect is at least a third less than that computed. It should be borne in mind, however, that Kirchhoff's formula applies primarily to electrodes in air. To completely simulate this condition with a solid dielectric would require that the electrodes be completely surrounded by a considerable thickness of the dielectric. If the difference noted above is due to lines of force which pass partly through air and partly through the sample, this difference should be greatest for a sample of high dielectric constant and diminish as the dielectric constant of the sample approaches that of the air. It is seen from Fig. 3 that this is the case, the curve for hard rubber having a dielectric constant of 3 being nearer to the computed curve than that for glass having a dielectric constant of 7.7. The anomalous shape of the curve for phenol fibre is apparently due to a non-uniformity in the samples which will be discussed later.

That the above results are materially affected by flux passing partly through air and partly through the sample was proved by an additional test. The 3/8" phenol fibre sample with 4½" discs on each side

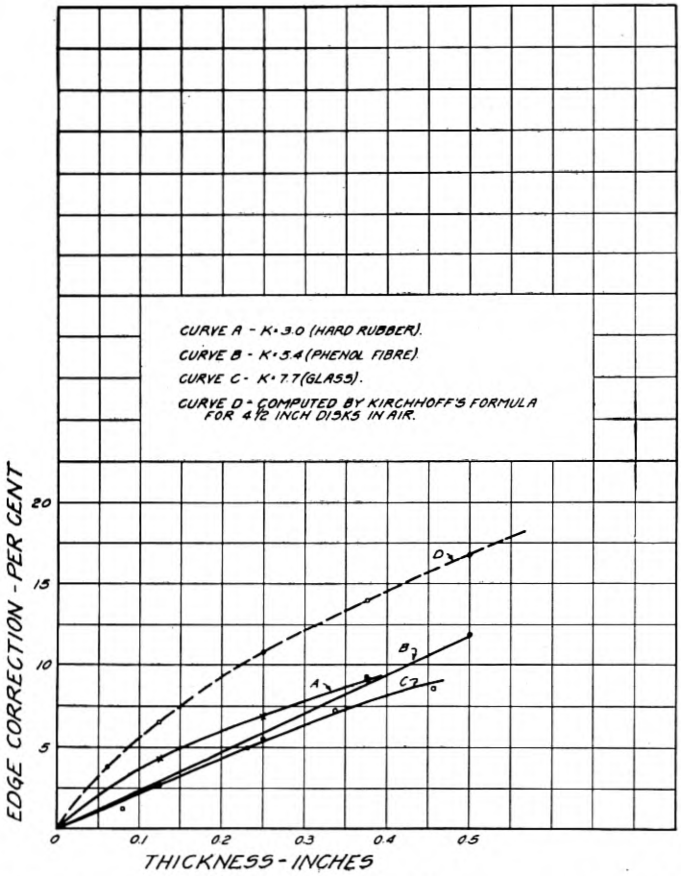


Fig. 3—Percentage edge correction for samples 6 inches square with foil electrodes both 4½ inches in diameter

was first measured in air. Since the upper electrode could not be entirely covered and still obtain contact with it, the ¼" phenol fibre plate was laid on top of it so as to cover just one-half of the sample and electrode. This caused an increase of 1.4 mmf. Using the ¼" plate in the same way to cover one-half of the lower surface of the sample and electrode, the increase was .7 mmf. showing that when one electrode is grounded the effect is not symmetrical. Covering both sides of the sample should, therefore, cause an increase of 4.2 mmf. or 8.6% of the capacitance of the sample. Adding this to the 9.3% fringe effect already determined makes a total of 17.9% as compared with 13.9% by Kirchhoff's formula. At least part of this difference is also due to the non-uniformity referred to above which is very marked in this sample.

A similar test was made on the ⅛″ phenol fibre sample. However, this sample was somewhat warped so that there was an appreciable air-gap between the electrode and the cover plate. The total effect computed as above was 2.6%, which added to the 2.8% already determined makes a total of 5.4% or slightly under the value computed by Kirchhoff's formula. This, no doubt, is accounted for by the air-gap.

The agreement with Kirchhoff's formula is reasonably good, therefore, for disc electrodes completely surrounded by dielectric, but it is evident that the formula does not apply to the case of disc electrodes applied to sheet materials.

The fringe effect for the 4½″ upper circle and 6″ lower square electrodes is shown in Fig. 4. It is found to be about 2½ times as large as for the equal 4½″ electrodes. It also varies somewhat with the dielectric constant of the sample, being greater for the lower dielectric constant. The anomalous behavior of the phenol fibre samples is shown in this figure also.

From Item 8 of Table II, it is seen that when a shallow metal ring is placed on the upper disk to simulate the conditions of a mercury electrode, a further increase of from 2 to 4% of the true capacitance of the sample takes place. This likewise is greater the thicker the sample and the lower its dielectric constant. A similar test for the increase in capacitance due to vertical height of the metal ring was made for the more exaggerated case of a 4″ disk and a ring ¾″ high; the lower electrode being 6″ square. In this case the increase varies from 2½% for ⅛″ phenol fibre to 8% for ⅜″ hard rubber. This shows the importance of keeping the vertical dimension of the metal ring as small as possible.

Insulating Ring

In order to determine the corresponding effect when an insulating ring is used for confining mercury, a somewhat similar test was made. Several different rings were used as follows: ring No. 1 is ¾″ high cut from hard rubber tubing having a ⅛″ wall with the edges cut square. Ring No. 2 was the same as above except that the edge was beveled on the outside as in Fig. 2-A to a thickness of 1/64″. Ring No. 3 was cut from phenol fibre sheet 1/8″ thick and had a radial width of 3/8″. Since rubber tubing of the desired size was not available the rubber rings were somewhat smaller than 4½″ in diameter but in all cases the foil electrode with which they were used was cut

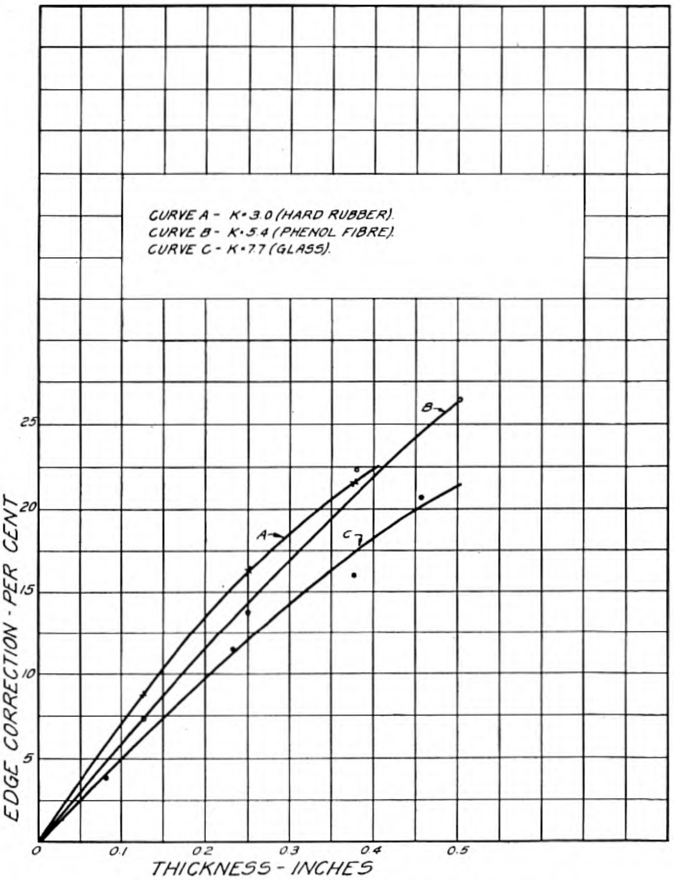


Fig. 4—Percentage edge correction for samples 6 inches square with foil electrodes upper, 4½ inches diameter, lower, 6 inches square

to exactly the same size as the inside diameter of the ring and the results were reduced to the equivalent value for 4½″ diameter.

In this test in order to simulate the depth of the mercury without changing the electrode contacts, the lower part of the inner surface of the ring was coated with foil to a height equal to the assumed depth of the mercury. This was taken as 1/8″ in each case which is about the minimum depth of mercury that can be depended upon to cover the surface of the sample and fill the angle between the sample and the ring. This, of course, has slightly less effect than if the entire surface of the electrode were raised as is the case when mercury is poured into the ring, but the difference is probably negligible. The procedure was the same as previously described for the metal ring. The lower foil electrode covered the entire surface of the sample. The upper foil is the same diameter as the inside of the insulating ring. (In the case of the metal ring the foil was the same size as the outside diameter of the ring at the lower edge). The capacitance is measured first between the foils alone and then with the ring superimposed on the upper foil. The change should represent very closely the increased edge effect due to a depth of 1/8″ of mercury in the ring.

The results of this test are tabulated in Table III. It will be noticed that the change in capacitance due to the beveled rubber ring is about the same as for the metal ring, while the change due to the square edged ring is materially greater than for the metal ring. The flat phenol fibre ring produces a change in capacitance two or three times as great as the metal ring and the apparent power factor for the rubber sample is more than doubled. This, of course, is due to the dielectric loss in the ring itself and therefore the lower the power factor of the sample under test the greater is the proportional error. While the rubber rings had no appreciable effect on the power factor of the samples tested, it is believed that in the case of very low loss materials such as fused quartz the effect might still be appreciable. Since the insulating rings under the best conditions are no better than a metal ring and under poor conditions are very much inferior, it is believed that in general metal rings will be found more satisfactory for confining mercury electrodes.

Thus far we have considered the experimental data primarily with regard to the capacitance and dielectric constant. Table IV shows the power factors computed from the conductance readings corresponding to readings 1, 4, 6 and 10 in Table I. These values were computed from the capacitance and conductance values for the various types of electrodes and without correction for edge effects or ground capacitances. The values for hard rubber illustrate fairly well the variation for different electrodes which would be expected on the basis of the preliminary discussion. The value obtained with the 4½″ circle and guard ring should be the true value. The other values should be slightly lower on account of the additional capacitance without corresponding power less due to the flux which passes partly and entirely through air. In general these variations are small and almost beyond the limit of accuracy of the measurements. However, a careful analysis of the results (omitting the 3/8″ phenol fibre sample which will be discussed separately) seems to indicate that the power factor values obtained with the two 4½″ circles agree slightly better with those obtained with the guard ring electrode than do those obtained with any of the other electrodes.

In the case of the 3/8″ phenol fibre sample the variations with different electrodes are much greater than the probable inaccuracy of the measurements. Apparently, they can only be attributed to non-uniformity of the material in different parts of the sample. Therefore, a special test to determine this fact was made on this sample. By interchanging the connections to the guard ring and the 4½″ center electrode, it is possible to measure the capacitance and conductance of the outer part of the sample without including the center part. The sum of the values of the two parts checks well with the value for the entire 6″ square. These results show that while the inner part of the sample has a power factor of 1.97%, the power factor of the outer part is 3.17% or approximately 60% higher. The corresponding dielectric constants are 5.08 and 5.48, respectively. The reason for this difference probably is that since the material is of a laminated nature moisture penetrates more readily from the edges than from the sides of the sample and thus causes a progressive variation of the electrical characteristics from the edges to the center of the sample. It is obvious that when the two 4½″ electrodes are used without guard rings, some of the outer part of the sample is included due to the fringe effect and that when the 4½″ upper and 6″ lower electrodes are used still more of the outer part is included for the same reason, and the values obtained are increased accordingly.

As previously mentioned, it is probable that this non-uniformity is responsible for the different shape of the edge effect curve for phenol fibre as compared with those of hard rubber and glass and it is almost certainly the cause of the point representing this particular sample being exceptionally high. Since there are wide variations in the power factors of the different samples of both phenol fibre and glass it is possible that there are minor variations through the sample due to causes other than moisture and that these may account for some of the other apparent irregularities in the results.

Method of Applying Corrections

While percentage values are the most convenient for discussion of the relative importance of the various corrections involved in the use of a given type of electrode, the absolute values of these corrections in micro-microfarads are probably somewhat more convenient for actual use in making the necessary calculations.

Consider the case of mercury electrodes with the lower electrode grounded. On the basis of the foregoing discussion the total capacitance C which is measured may be considered as made up of four parts, namely, C_x the capacitance between the electrodes which would exist under uniform field conditions as when a guard ring is used. C_{e1} the edge effect which would exist if the upper electrode were a thin disk. C_{e2} the additional edge effect due to the height of the metal ring. C_g the capacitance to ground of the upper surface of the electrode. The dielectric constant $K = 4.46 \dfrac{C_x d}{A}$ where C_x is in micro-microfarads d and A are the thickness and area of the sample in inches and square inches, respectively. To compute K, C_x must therefore be obtained from the relation

$$C_x = C - (C_{e1} + C_{e2} + C_g).$$

Values for C_{e1}, C_{e2}, C_g are given in the tables for certain particular cases and may be determined in a similar manner for any other cases. For foil electrodes, conditions are similar except that C_{e2} is zero. C_{e1} which is the largest of the corrections is plotted in Fig. 5 for various values of K and several thicknesses of the sample as taken from the previous tables. It will be seen that in order to apply this correction the approximate value of the dielectric constant of the sample must be known in advance. This can always be obtained by making a preliminary computation neglecting the corrections entirely.

As an example of the above method suppose measurements have been made on a 1/8″ sample of material having a dielectric constant of about 4.5 using mercury electrodes with a shallow metal ring. From Curve A, Fig. 5 we get 9.2 mmf. for C_{e1}. C_{e2} is estimated by interpolation from item 7 of Table II at about 2.7 mmf. C_g is taken as the average for item 6 of Table II or 0.8 mmf. This makes a total of 12.7 mmf. to be subtracted from the measured capacitance in order to obtain C_x from which the dielectric constant is computed. If 4½″ foil electrodes were used C_{e1} would be taken from Curve B, Fig. 5 as 4.0 mmf. and C_g from the average of item 11, Table II as 2.4 mmf. making a total correction of 6.4 mmf.

The values given in Fig. 5 for C_{e1} are, of course, applicable only to a given size of electrode, namely 4½″ in diameter. If the edge effect capacitance is considered equivalent to that of an additional ring electrode surrounding the main electrode, this capacitance would be proportional to the mean radius of this ring, or to the radius of the upper electrode plus one-half of the width of the ring. If this equivalent ring is of constant width for various sizes of electrodes, the edge correction would be proportional to the radius of the electrode plus a constant. As Fig. 5 shows that the correction in micro-microfarads does not vary greatly with the thickness of the sample, the width of

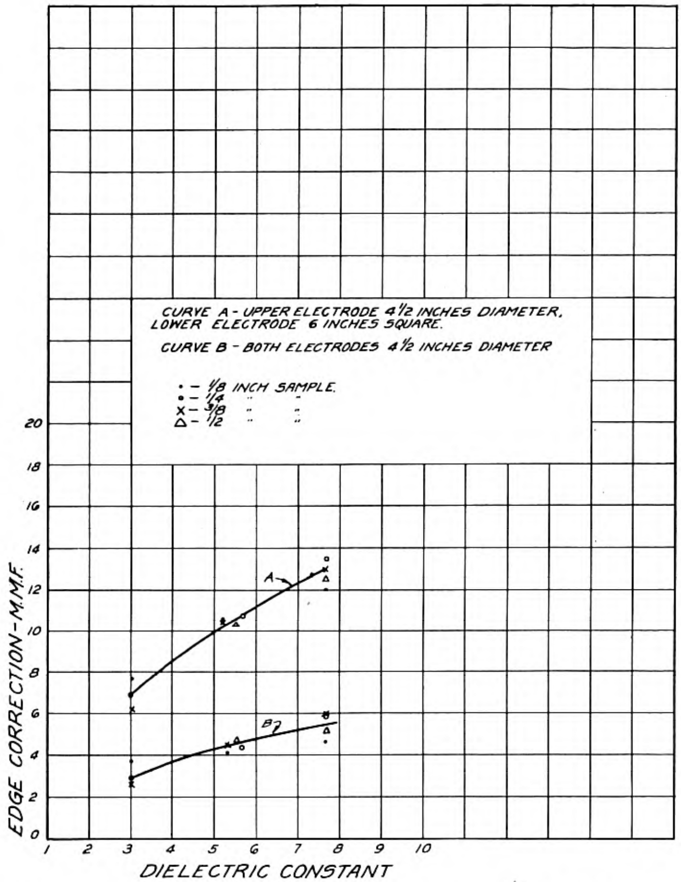


Fig. 5—Edge correction versus dielectric constant for samples 6 inches square with foil electrodes

the above hypothetical ring must be approximately proportional to the thickness of the sample. Computations for several samples show that the width of this ring is of the order of twice the thickness of the sample. Hence it appears that as a first approximation the edge correction for various sizes of electrodes may be taken as proportional to $r+t$, r being the radius of the electrode and t the thickness of the sample. Table V shows a series of measurements on one sample with electrodes of several different sizes. In this table the edge correction C has been taken as the difference between the capacitances measured with and without a guard ring. This includes the direct capacitance of the upper electrode to ground. It will be seen that while the value of $\frac{C}{r+t}$ is not exactly constant it varies much less than $\frac{C}{r}$. Hence if C_1 is the edge correction for an electrode of radius r_1 the edge correction for an electrode of radius r_2 under similar conditions would be approximately

$$C_2 = C_1 \frac{r_2+t}{r_1+t}.$$

This, of course, applies only to the case of one large and one smaller electrode. For the case of the equal electrodes t in the above expression probably should be multiplied by a constant of the order of 0.4.

Tables for the Article

By E. T. Hoch

TABLE 1

Capacitance Readings as Described under "Evaluation of Errors" Micromicrofarads

Reading No.	Phenol Fibre				Hard Rubber			Glass			
	½ in.	⅜ in.	¼ in.	⅛ in.	⅜ in.	¼ in.	⅛ in.	.455 in.	.338 in.	.232 in.	.080 in.
1	94.7	116.3	*	335.0	69.5	100.5	200.4	147.0	190.2	278.0	790.8
2	*	113.1	*	332.2	66.7	97.8	195.1	*	*	*	*
3	*	3.2	*	3.3	*	3.3	3.3	*	*	*	*
4	39.7	48.6	60.2	146.6	28.9	42.3	86.4	61.4	82.0	118.4	344.6
5	(50.1)	59.1	91.0	157.2	35.1	49.1	94.1	(74.0)	(95.0)	(131.9)	(356.6)
6	50.8	59.6	92.4	157.7	35.7	49.4	94.4	74.7	95.7	132.6	357.3
7	*	60.9	*	160.8	36.9	51.3	96.5	*	*	*	*
8	*	0.8	1.3	1.0	0.7	0.6	0.7	*	*	*	*
9	(44.7)	53.1	84.6	150.7	31.5	45.2	90.1	(66.6)	(88.0)	(124.3)	(349.2)
10	46.7	55.7	87.9	153.1	33.6	47.2	91.2	66.9	90.3	126.6	351.5
11	*	2.6	3.0	2.3	2.3	2.3	2.1	*	*	*	*

* Not measured.

NOTE: Readings 5 and 9 given in parentheses are derived from readings 6 and 10 respectively using average values for readings 8 and 11.

TABLE 2

Values Computed from Table 1 and Comparison with Measured Values

Value Computed (mmf. unless otherwise stated)	Phenol Fibre				Hard Rubber			Glass			
	½ in.	⅜ in.	¼ in.	⅛ in.	⅜ in.	¼ in.	⅛ in.	.455 in.	.338 in.	.232 in.	.080 in.
(1) Cap. of 6 in. square electrode to ground—Reading 1–2 (Table 1)................	*	3.2	*	2.8	2.8	2.7	5.3				
(2) Ditto by direct meas. Reading 3....................	*	3.2	*	3.3	*	3.3	3.3				
(3) Fringe effect 4½ in. circle to 6 in. square—Reading 5–4...	10.4	10.5	10.8	10.6	6.2	6.8	7.7	12.6	13.0	13.5	12.0
(4) Ditto in per cent..........	26.2	22.4	13.5	7.2	21.4	16.1	8.8	20.5	15.8	11.4	3.4
(5) Stray cap. to grd. from 4½ in. upper with 6 in. lower electrode (6–5)............	*	0.5	1.4	0.5	0.6	0.3	0.3				
(6) Ditto by direct meas. Reading 8....................	*	0.8	1.3	1.0	0.7	0.6	0.7				
(7) Additional capacitance due to metal ring Reading 7–6...	*	1.3	*	3.1	1.2	1.9	2.1				
(8) Ditto in per cent..........	*	2.7	*	2.1	4.1	2.6	2.4				
(9) Fringe effect 4½ in. circle to 4½ in. circle—Reading 9–4..	4.7	4.5	4.4	4.1	2.6	2.9	3.7	5.2	6.0	5.9	4.6
(10) Ditto in per cent..........	11.8	9.3	5.5	2.8	9.0	6.8	4.3	8.4	7.3	5.0	1.3
(11) Stray cap. to grd. from 4½ in. upper with 4½ in. lower electrode (10–9)...........	*	2.6	3.3	2.4	2.1	2.0	1.1				
(12) Ditto by direct meas. Reading 11....................	*	2.6	3.0	2.3	2.3	2.3	2.1				
(13) Fringe effect by Kirchhoff's formula for 4½ in. circles in air (per cent)....					13.9	10.7	6.6				

*Not determined.

TABLE 3

Effect of Various Types of Rings for Confining Mercury Electrode

	Increase in Capacitance Per Cent		Power Factor Per Cent	
	1/8 in. Glass	3/8 in. Hard Rubber	1/8 in. Glass	3/8 in. Hard Rubber
Without ring	0.0	0.0	2.14	.43
With metal ring 3/16 in. high, outside bevel	2.1	4.1	2.11	.45
With square edged hard rubber ring	2.7	7.0	2.11	.45
With bevel edged hard rubber ring	1.8	4.9	2.13	.44
With flat phenol fibre ring	4.7	12.6	2.45	1.08

TABLE 4

Power Factors as Measured with Different Electrodes

Electrodes	Phenol Fibre				Hard Rubber			Glass			
	1/2 in.	3/8 in.	1/4 in.	1/8 in.	3/8 in.	1/4 in.	1/8 in.	.455 in.	.338 in.	.232 in.	.080 in.
6 in. squares	2.40	2.24	1.95	2.18	0.43	0.41	0.45	1.67	2.46	2.18	2.94
4½ in. circle with guard ring	2.56	1.96	2.05	2.19	0.44	0.47	0.44	1.81	2.43	2.21	2.86
4½ in. upper, 6 in. lower	2.50	2.16	2.08	2.20	0.40	0.39	0.46	2.00	2.43	2.14	2.85
Ditto with metal ring	*	2.12	*	2.18	0.39	0.39	0.46	*	*	*	*
4½ in. upper and lower	2.58	2.14	2.11	2.20	0.44	0.45	0.44	1.80	2.50	2.20	2.90

* Not measured.

TABLE 5

Measurements on 1/4 In. Phenol Fibre with Electrodes of Various Sizes

Radius of Upper Electrode In.	Cap. With Guard Ring Mmf.	Cap. Without Guard Ring Mmf.	Edge Correction C Mmf.	C/r	C/r+t	Diel. Const. Computed from 2nd Col.
.51	4.25	7.8	3.55	7.0	4.7	5.7
1.00	15.9	21.9	6.0	6.0	4.8	5.58
1.50	35.4	44.1	8.7	5.8	5.0	5.50
2.00	62.6	73.2	10.4	5.2	4.6	5.50
2.26	82.1	92.9	10.8	4.8	4.3	5.60

NOTE: Sample was 6 in. square with lower electrode covering entire lower surface.

Load Carrying Capacity of Amplifiers

By F. C. WILLIS and L. E. MELHUISH

SYNOPSIS: This paper describes the adaptation of the cathode ray oscillograph to the determination of the overload point of vacuum tube amplifiers. Using the input voltage to produce a horizontal deflection, and the output voltage or current to produce a vertical deflection, the amplifier performance is readily determined by noting the resulting figure on the fluorescent screen. So long as the figure is virtually a straight line or an obviously undistorted ellipse, it was found that the amplifier output is free from harmonics. As soon as overloading begins, the oscillogram shows either a sharp bend at either or both extremities of the line or apparent distortion of the ellipse. The method has the advantage of being quick.

IN any device used for the amplification of a complex electrical wave such as that necessary for the transmission of speech or music, distortion may arise in two ways. (a) The amplification may not be the same for all frequencies in the band to be transmitted. (b) The relationship between input voltage and output current may not be such as can be described by a straight line when r.m.s. values are plotted against each other.

Considering the distortion due to cause (b) alone it is true that for most practical devices the relationship between input and output can be described by a curve which is approximately straight for a portion of its length but as the amplitude of the wave to be transmitted increases, operation is over a longer portion of the curve and it ceases to be possible to regard the characteristic curve as straight. It therefore becomes necessary to determine for any design the maximum energy that the system can carry without noticeable distortion from this cause. For a system intended to transmit speech or music the final decision as to how much distortion is permissible must depend upon the judgment of the expert listener but analytical methods are of service in establishing reference points by measurements which can be duplicated without reference to any particular person. The purpose of this paper is to describe some work undertaken for this purpose.

Departure from the ideal straight line relationship between input and output results in the production of harmonic overtones of all the frequencies present in the input and in the production of beat notes between frequencies if there is more than a single frequency in the input. It follows from this that if a complex wave is passed through a device having a characteristic of this nature the output will differ from the input in the proportion of the different frequencies and may contain frequencies that were not present in the input at all. A change in the proportion of the different frequencies may be partly due to cause (a) but the introduction of new frequencies can only be due to cause (b). As is now known (BELL SYSTEM TECHNICAL JOURNAL October, 1923) the response of the ear itself is non-linear so that subjective harmonics and sum and difference tones are heard by every listener. Under good conditions, therefore, the distortion produced by the non-linear transmission of the amplifier will be so small compared to that produced in the ear itself that it will not be noticed. On the other hand, it may be so great as to render unrecognizable the speech or music being transmitted. This condition will be familiar to everyone who has been compelled by an enthusiastic friend to listen to a heavily overloaded radio receiver.

The principal parts of a vacuum tube amplifier where one might expect to find the non-linear response under consideration are in the magnetic circuits of the transformers and retard coils and in the vacuum tubes. Generally, in the amplifiers considered in this paper, the design of the transformers is such that the magnetic flux density is small so that the magnetization curve is practically a straight line and very little distortion is to be expected from this cause. On the other hand the $E_c - I_b$ characteristic of the vacuum tube is approximately straight for only a small portion of its length and has a pronounced curvature in the usual working range. This is the case even for well designed circuits where under proper operation there is no possibility of the grid of the tube drawing current. It is therefore, in the characteristics of the vacuum tubes that the principal source of trouble of this nature is to be looked for. That this anticipation is justified will be shown by the results described in this paper.

The relationship between output and input of a vacuum tube has been studied from a mathematical viewpoint and formulae have been established by which the resultant output for a given input may be calculated provided the tube parameters and circuit impedances are precisely known. For any commercial amplifier having several stages the measurement of these quantities and the necessary calculations would be a slow procedure. By experimental methods it is possible to determine directly and quantitatively the distortion that occurs in any particular case. This has been done for a number of amplifiers under various load conditions with a view to establishing convenient criteria by which it is possible to determine quickly and easily how much energy any amplifier will transmit without serious distortion. The amplifiers dealt with were all audio-frequency amplifiers so that no questions of radio frequency amplification or of intentional rectification or modulation are considered and the measurements were all made with single frequency inputs as this naturally forms the basis for a more complete analysis of the problem. Three kinds of measurements were made.

1. The gain of the amplifier was measured under load conditions which varied from a point well below its carrying capacity to a point where it was obviously overloaded. The results from tests of this kind show that the gain is uniform at low outputs, begins to fall off when the output reaches a certain level and falls off more and more rapidly as the load is further increased. The point at which the gain begins to fall off has sometimes been taken as a criterion of the load carrying capacity of the amplifier.
2. With a single frequency (1,000 c.p.s.) input to the amplifier the output was analyzed at a number of points along the load-gain curve and the percentage of harmonic to fundamental in the output plotted against the same scale of energy output as for the load-gain curve.
3. The input voltage was made to produce a horizontal deflection in a cathode ray oscillograph while the output voltage or current was made to produce a vertical deflection. In effect this is a convenient means of drawing the input-output characteristic of the amplifier so that its general curvature and the loads at which any sudden changes of curvature occur may be easily observed. For an amplifier that produced no distortion or phase shift, the resultant figure would be a straight line whatever the wave shape of the input. If there were phase shift but no distortion the result would be an ellipse or circle depending on the phase and amplitude relationships of the input and output provided the input were a pure frequency. In general for the practical case the result is a distorted ellipse showing that the wave undergoes both distortion and change of phase in passing through the amplifier. With increasing load the distortion becomes more and more pronounced.

The gain measurements were made by methods in principle the same as those embodied in standard gain measuring sets of the Bell System, and while it is not the intention of this paper to give a detailed description it may be desirable to give a brief statement of the principles involved. For the purposes of this paper the gain of an amplifier is defined as the logarithm of the ratio of the power delivered into its load impedance to the power that would be delivered if the amplifier were removed and replaced by the best possible passive network.

Thus

$$N_{TU}=10 \text{ Log}_{10}\left(\frac{W_o}{W_i}\right).$$

In practice an amplifier is almost always measured between impedances that are pure resistances. These impedances are set up by variable resistance networks so designed that when the current in one mesh of the input network is equal to the current in a mesh of the output network the gain of the amplifier may be read from the settings of the dials and switches controlling the networks. An indicating device which may be switched from the input network to the output network indicates when the currents in the two meshes mentioned are equal. Where sufficient energy is available a thermo-

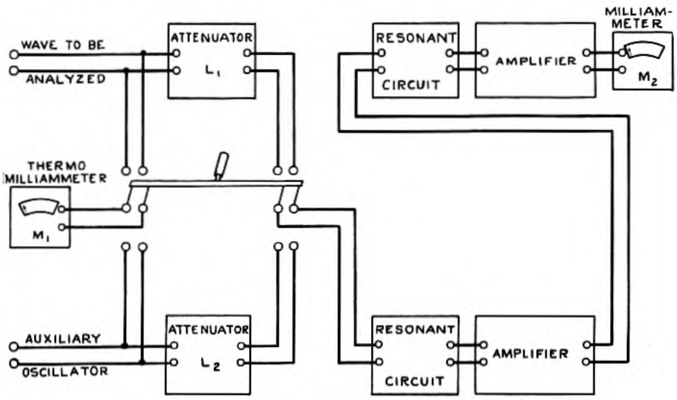


Fig. 1—Circuit for harmonic analysis

couple is used as the indicating device. It follows that measurements are then always in terms of r.m.s. values of the currents employed. So long as these currents are approximately single frequency the r.m.s. value of the whole wave is very nearly equal to the r.m.s. value of the fundamental. In all the experiments described here gain measurements were made by thermocouple and as will be seen later the proportions of harmonics were such that unless the amplifier was overloaded the discrepancy between the r.m.s. measurements made and those that might be made by methods taking account of the fundamental only was usually less than 1%, and in the cases of extreme overload less than 5%.

The harmonic analyses were made by an electrical analyzer whose principles of operation are indicated in Fig. 1.[1] In operating this

[1] The principles of this analyzer are fully described in a paper by Mr. A. G. Landeen to be published in the *B. S. T. J.*

analyzer the resonant circuits of the amplifier detector and the auxiliary oscillator were tuned to the harmonic to be measured. The input currents and the attenuators L_1 and L_2 were adjusted so that the readings of meters M_1 and M_2 did not change when the four-pole, double-throw switch was thrown from one position to the other. It will be seen that the difference between the settings L_1 and L_2 then gives the proportion of harmonic to total r.m.s. value of the output wave of the amplifier. The proportion of harmonic necessary to cause a 1% change in the r.m.s. value of a wave is approximately 14% and for the values obtained in the experiment the r.m.s. value of the output could be taken as equal to the fundamental in the output. The difference between L_1 and L_2 therefore gives the proportion of harmonic to fundamental. In the present work the difference between the frequencies to be separated was large enough to avoid any difficulty in obtaining sufficient resolution by the use of simple resonant circuits.

It is, of course, necessary that for measurements of this kind the current supplied to the amplifier under test should be a single frequency. A vacuum tube oscillator which was known to give a very pure wave was used as the source of current. To obtain sufficient energy for all the measurements made it was necessary to amplify the output current from this oscillator and subsequently filter it to remove the harmonics introduced by the amplifier. Final analysis of the wave applied to the amplifier under test showed in most cases less than 0.2% and in all cases less than 0.5% of third harmonic and less than .1% of all other harmonics. Greater purity could have been obtained at the expense of more time and trouble but this was considered sufficient for the purposes in view. Where necessary a small correction for the harmonic content of the input wave has been applied to the results.

The voltages to operate the cathode ray oscillograph were obtained by a step-up transformer for the input and directly off a resistance potentiometer for the output. The use of a step-up transformer for the output wave is in general undesirable because it introduces phase and amplitude changes which differ for the component frequencies of the wave and thus the transformer itself introduces a distortion which renders the interpretation of the figure as applied to the amplifier distortion more difficult. This limits the method to cases where a minimum of 10 volts is available in the output. For the amplifiers dealt with here this voltage was available and the limitation was not felt. On the input side the step-up transformer has to transmit one frequency only so that the same difficulty does not occur. Pictures of the figures obtained on the fluorescent screen were taken with an ordinary camera with from one to three-minute exposures. On most of the pictures the horizontal and vertical axes were also recorded on the screen.

The circuits of the amplifiers dealt with and the results of the analyses are graphically presented in the following figures. Amplifier

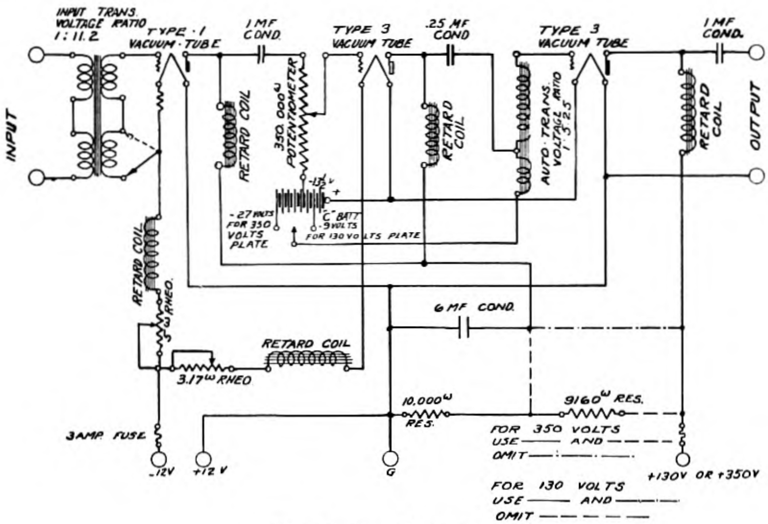


Fig. 2—Amplifier No. 1

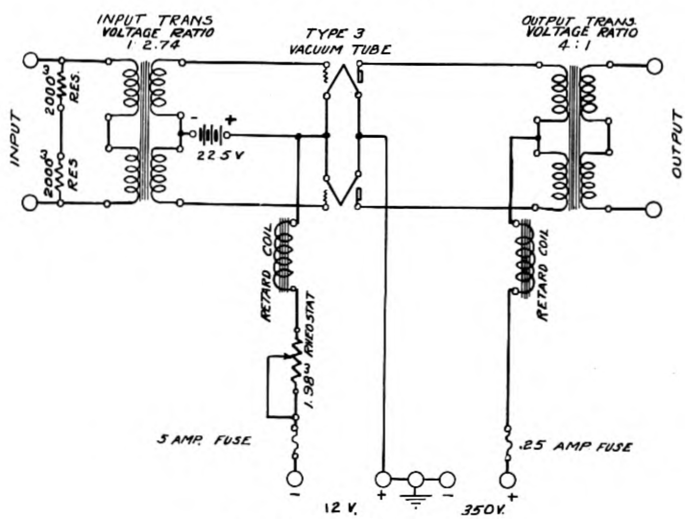


Fig. 3—Amplifier No. 2

No. 1 shown in Fig. 2 is used for amplification from very low up to medium powers in Public Address, Radio Broadcasting and similar systems. Provision is made for operating this amplifier on either 130 or 350 volt anode potential and tests were made under both these conditions. This amplifier is designed to work from an impedance of 200 ohms into one of 4,000 ohms. Amplifier No. 2 shown in Fig. 3

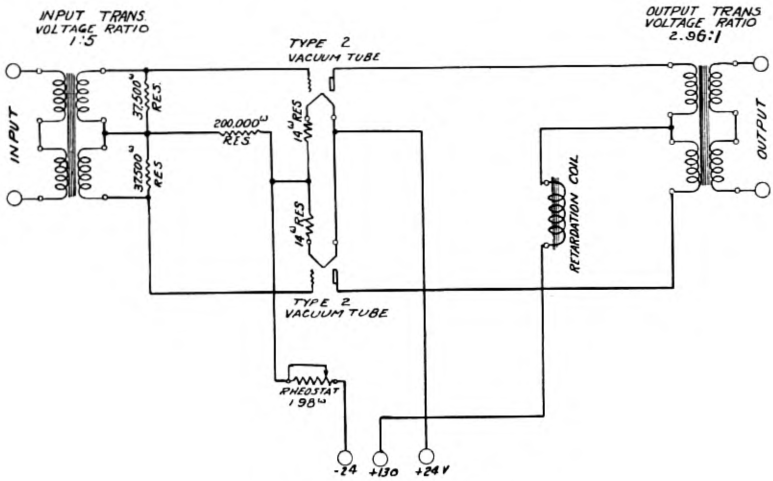


Fig. 4—Amplifier No. 3

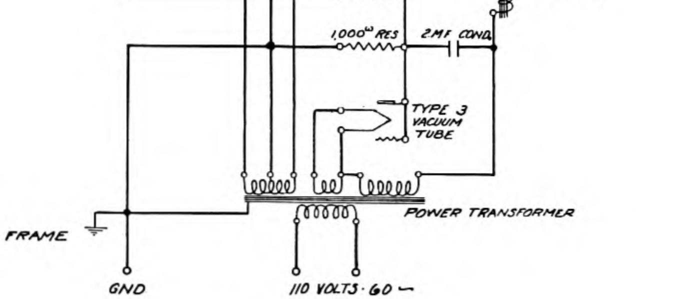


Fig. 5—Amplifier No. 4

follows amplifier No. 1 as the final output stage of a Public Address System and as shown consists of one stage with tubes in the push-pull arrangement. To match amplifier No. 1 its input circuit is designed for 4,000 ohms and its output is stepped down by a transformer to 500 ohms. Amplifier No. 3 shown in Fig. 4, also a push-pull amplifier, is designed as the output stage of an oscillator used in field measurements in the telephone plant. Its input and output impedances are 2,000 ohms and 600 ohms, respectively. Amplifier No. 4 is intended for use in radio reception, amplifying the signals to loud speaker volume and operating from 110 volt, 60 cycle A.C. supply. As shown there is one rectifying tube for converting the 60 cycle a.c. to d.c. for supplying the anode of the amplifying tube. Its rated input and output impedances are 20,000 ohms and 4,000 ohms, respectively.

As will be noted from the diagrams there are employed in these amplifiers three types of vacuum tubes. The normal operating voltages and average characteristics of these tubes are shown in Table I.

TABLE I

	Filament Current	Anode Volts	Grid Biasing Volts	Anode Current Milli-amps.	Amplification Factor μ	Anode-Filament Impedance Ohms.
Vacuum Tube Type 1	1.00	130	− 1.6	0.7	30	60,000
Vacuum Tube Type 2	1.00	130	−20	25.0	2.5	2,000
Vacuum Tube Type 3	1.6	350	−27	25.0	6.5	4,000
Vacuum Tube Type 3	1.6	130	− 9	5.0	6.5	8,000

The harmonic analyses and load-gain curves of the amplifiers under the conditions noted are shown in Figs. 6, 7, 8, 9 and 10. The gain in transmission units and percentage of each harmonic up to the 5th together with the root of the sum of the squares of these percentages being plotted against watts output on a logarithmic scale. The oscillograph pictures taken at various points along these curves are shown in Figs. 11, 12, 13, 14 and 15.

As stated above, the oscillograph figure presents the input-output curve of the amplifier. Furthermore, if there is no distortion in the transformers, the horizontal deflection will be proportional to the alternating grid voltage applied to the first stage while the vertical deflection will be proportional to the alternating component of the plate current in the last stage. The figure drawn is then the dynamic $E_c - I_b$ charactertistic of all the tubes combined. That this is substantially the case may be seen by inspection of the figures. In the first oscillogram of Fig. 11 there is shown the characteristic for amplifier No. 1 under 130 volts plate supply and .047 watts output where the amplitude of grid voltage is such that no grid becomes positive with respect to the filament nor is the plate current reduced to zero at any part of the cycle. Under these conditions the tube characteristic as shown in the oscillogram has the same nearly parabolic shape as is found by other methods. The analysis made at

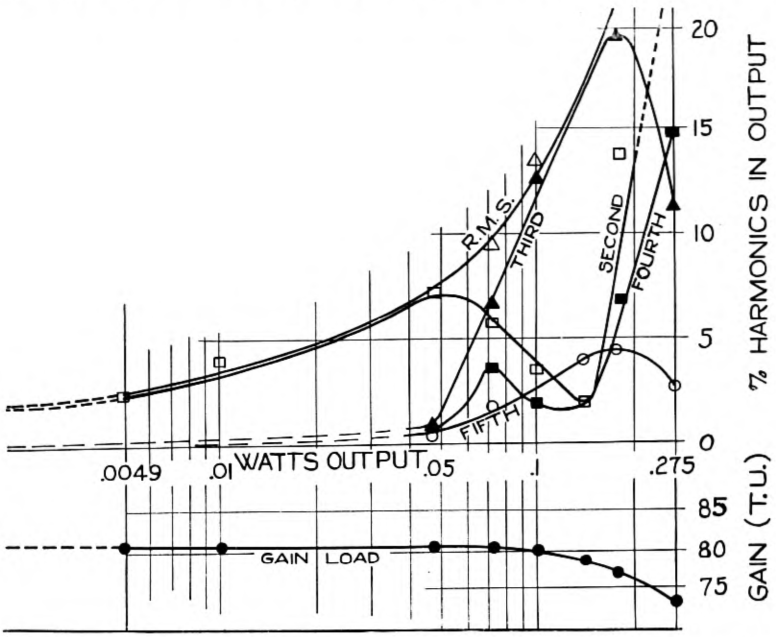


Fig. 6—Amplifier No. 1—Plate voltage 130-V, 1,000 c.p.s. input. Variation of gain and distortion with output

this point showed 7.3% second harmonic, .8% third and less than .1% fourth and fifth harmonics. The mathematical analysis of the problem expresses the $E_c - I_b$ characteristic of the tube by a power series and shows that the coefficient of the second power term in the series is the principal factor in producing second harmonic. The percentages of harmonic given therefore are such as would be obtained from a tube having a nearly parabolic characteristic.

Analyses made at lower outputs showed that the amount of second harmonic present varies with the power output in a manner described by a slightly curved line on the logarithmic scale used (Fig. 6). The third and higher harmonics are negligible at low outputs but at

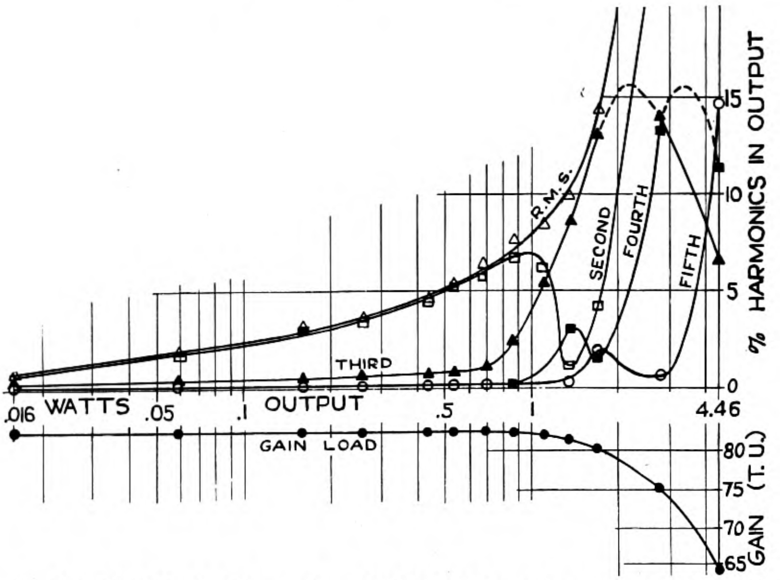


Fig. 7—Amplifier No. 1—Plate voltage 350-V, 1,000 c.p.s. input. Variation of gain and distortion with output

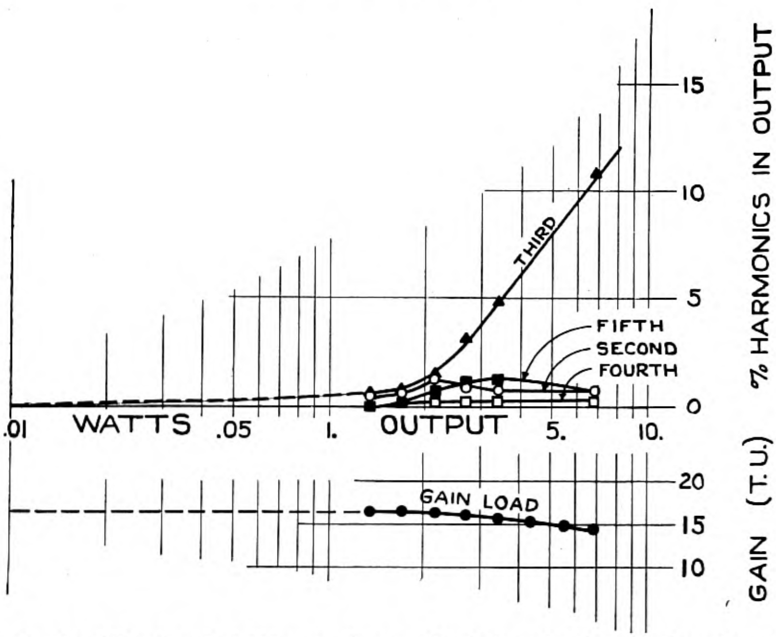


Fig. 8—Amplifier No. 2, 1,000 c.p.s. input. Variation of gain and distortion with output

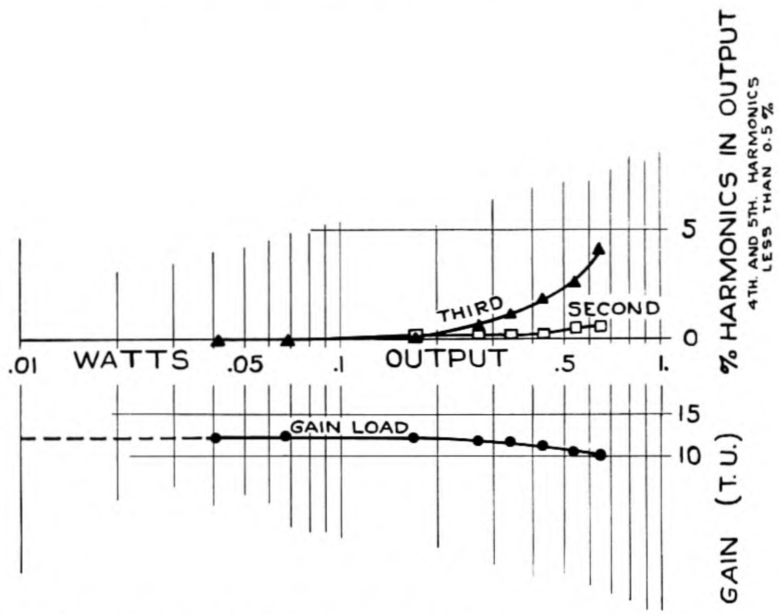


Fig. 9—Amplifier No. 3, 1,000 c.p.s. input. Variation of gain and distortion with output

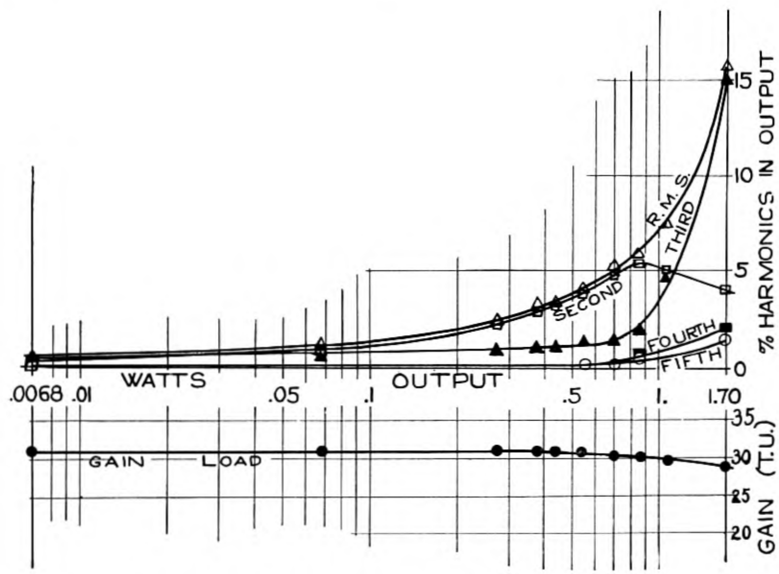


Fig. 10—Amplifier No. 4, 1,000 c.ps. input. Variation of gain and distortion with output

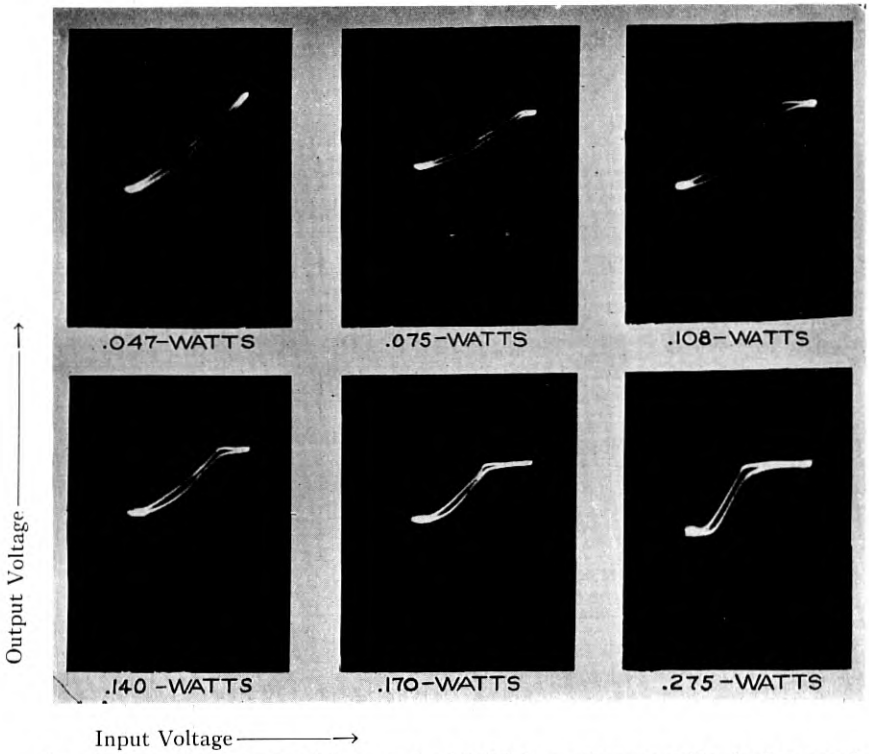


Fig. 11—Amplifier No. 1; 1,000 c.p.s.; load 4,000 ohm resistance, 130 volt plate supply

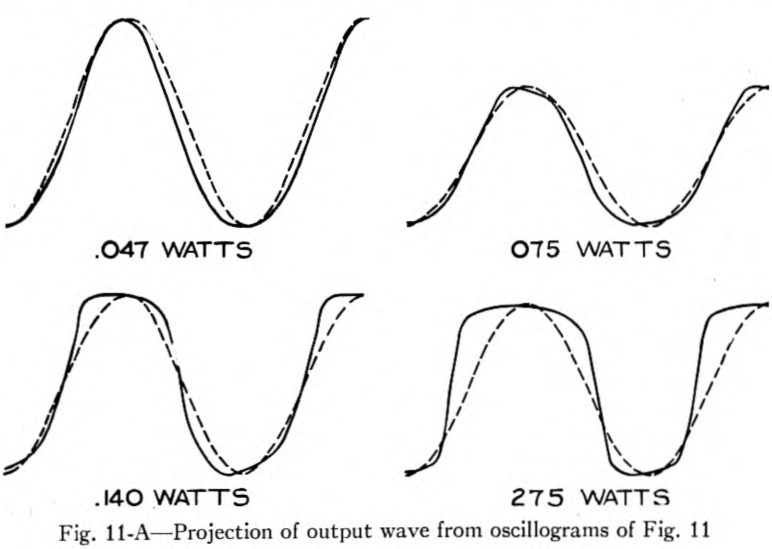


Fig. 11-A—Projection of output wave from oscillograms of Fig. 11

outputs above .05 watts the third harmonic increases rapidly and then falls somewhat while the second harmonic falls to a minimum of about 2% at .14 watt output and then increases rapidly. The fourth and fifth harmonics follow similar cycles of increase and decrease, the fourth following the second and the fifth the third.

To assist in interpreting the stages in the tube overloading at which these changes in the percentages of the different harmonics

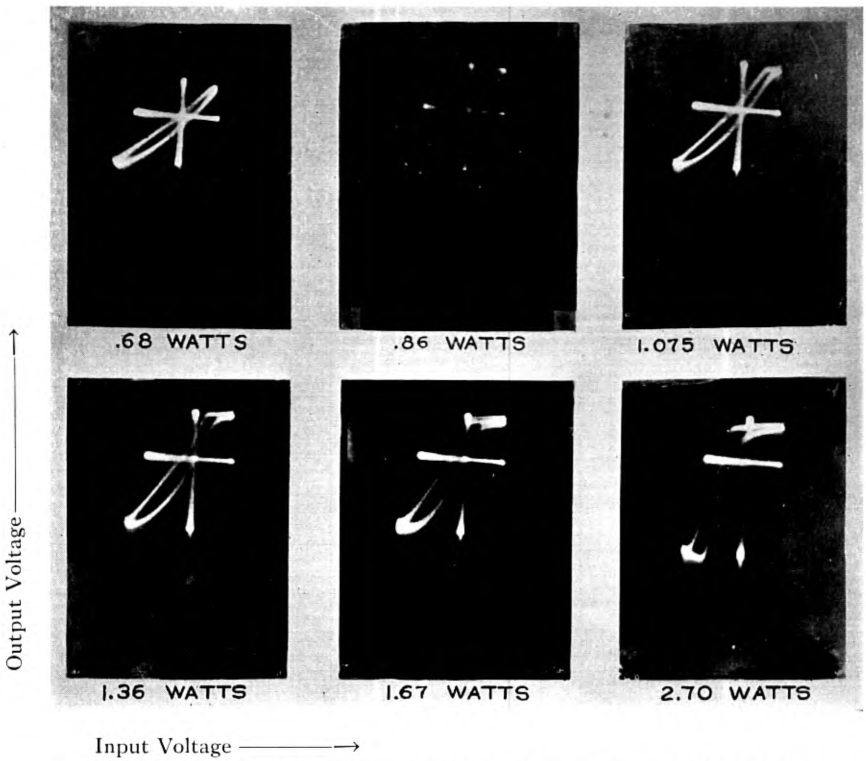


Fig. 12—Amplifier No. 1; 1,000 c.p.s.; load 4,000 ohm resistance, 350 volt plate supply

take place the waves corresponding to the vertical deflections of some of the oscillograms have been projected against a time axis by an appropriate geometrical construction, assuming the horizontal deflection a sine wave as it very nearly was.

These projections for four of the oscillograms of Fig. 11 are shown in Fig. 11-A, a pure sine wave being drawn against each figure for purposes of comparison. Up to an output of .047 watt the curved characteristic of the tube results in the asymmetrical wave shown with a predominant second harmonic. At a point slightly above .047 watt one or more of the grids becomes positive to the filament for part of a cycle and draws current. On account of the high impedance of the circuits supplying the grids this immediately results in a flattening of the top of the wave which is well developed at .075 watt output. This flattening of the top of the wave compensates to some extent for the curvature of the lower part of the tube characteristic so that the curve becomes more symmetrical with regard to the zero axis although more distorted. This corresponds to the fall of the second harmonic and increase of the third. At an output of .140 watt this compensation is more nearly complete than at any other output. At still higher outputs the top of the wave is still more flattened and the plate current is reduced to zero for a considerable part of the cycle so that the output wave becomes nearly rectangular as shown. This corresponds to large amounts of all harmonics.

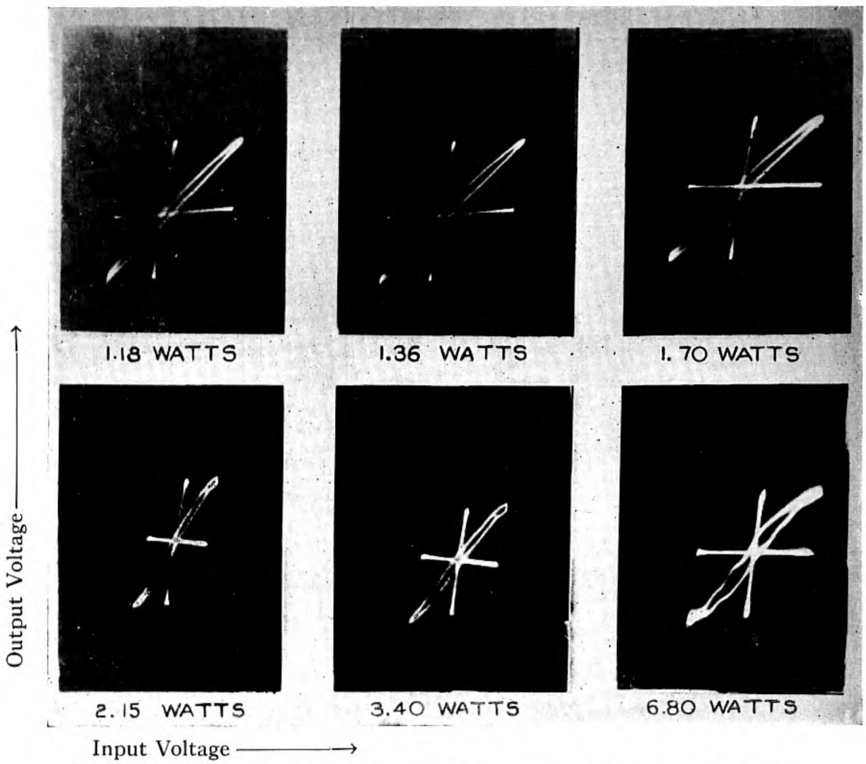


Fig. 13—Amplifier No. 2; 1,000 c.p.s., load 500 ohm resistance

In selecting from these results some point to be taken as the maximum carrying capacity of this amplifier, the question arises as to how much the second harmonic which is present at practically all loads will be noticeable. This, of course, depends on the training of the ear of the observer and on the quality of the music being transmitted. Comparing the note obtained from a cone type loud speaker when the wave corresponding to an output of .043 watts was applied with

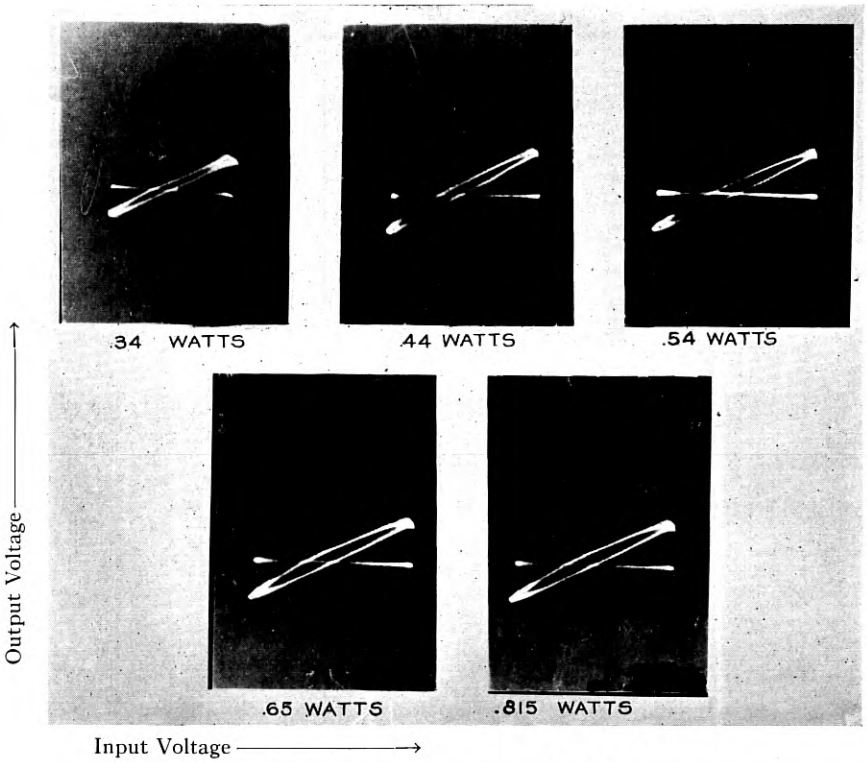


Fig. 14—Amplifier No. 3; 1,000 c.p.s.; load 600 ohm resistance

that obtained when the 1,000 cycle input was applied so as to produce a note of equal volume, it was found that it required a fairly sensitive ear to note the change in quality. On the other hand when a wave corresponding to the output wave obtained at a level of .068 watt was used for comparison the difference in quality between the pure tone and the distorted output was very easily noticed. From these data it may be assumed that when this amplifier is used in a system for the transmission and reproduction of speech or music it is fully loaded at an output of .04–.05 watt, representing the point where the grid of a tube begins to draw current and the third harmonic increases rapidly. It will be noted from the gain-load curve that the output has to increase to .1 watt before there is a noticeable falling off in the gain of the amplifier.

For the amplifier No. 1 under the condition of 350 volt plate supply the curves and oscillograms shown in Figs. 7 and 12 are of the same

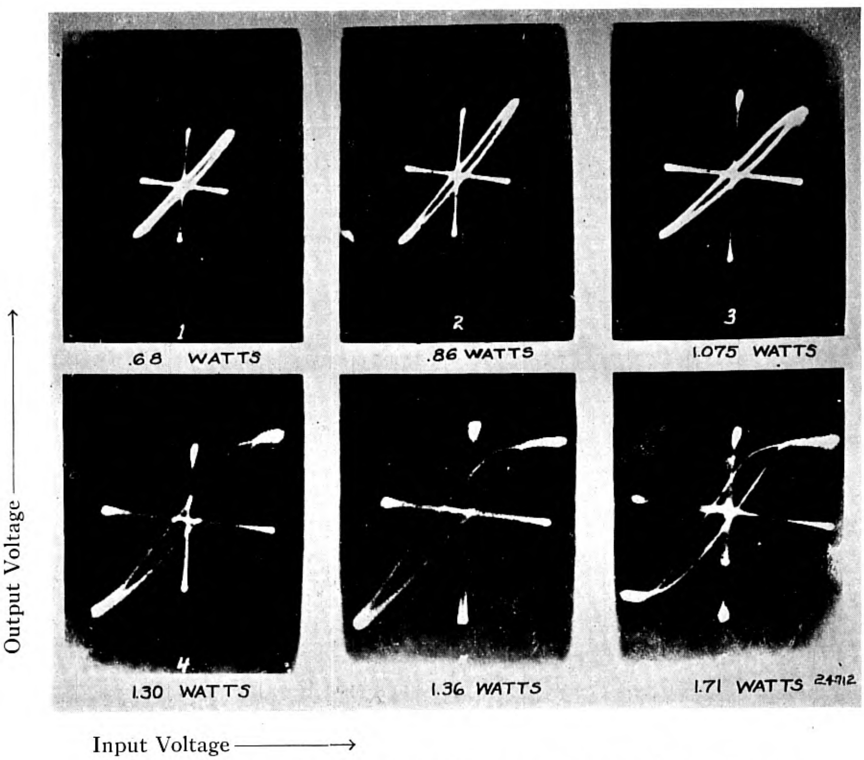


Fig. 15—Amplifier No. 4; 1,000 c.p.s.; load 4,000 ohms resistance

general nature. On account of the higher plate and grid biasing potentials an output of 0.7 watts is reached before the increase of third harmonic due to grid modulation occurs. This point is very definitely marked in the oscillograms for, of the pictures taken at outputs of 0.68 and 0.86 watts, the first is entirely free from this distortion, while the second where the current output is only 12.5% higher shows it clearly in the flattening at the top of the curve. On the other hand the gain-load curve does not show any appreciable decrease of gain until the amplifier is considerably overloaded.

For high power amplifiers and those amplifiers in which it is desired to reduce distortion to a minimum the push-pull arrangement of tubes has been used because with this arrangement the even harmonics generated in the tubes are suppressed in the output circuit. That the suppression is quite effective is shown by the curves and

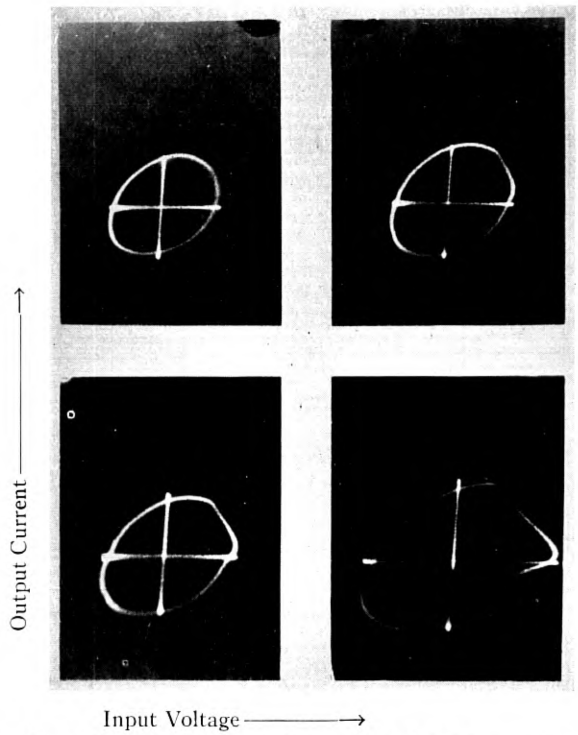


Fig. 16—Amplifier No. 4; 200 c.p.s. input; load 7,000 ohm negative reactance

oscillograms of Figs. 8, 9, 13 and 14 which were taken on the No. 2 and No. 3 amplifiers. These show that the even harmonics are very small at all loads and that the third harmonic increases suddenly at a point which in view of the plate and grid biasing potentials employed may be taken as the point at which grid modulation commences. On the oscillograms this point is not so clearly marked as in the previous cases but on those for No. 2 there is a slight flattening of the ends of the curve which is noticeable at 1.7 watts but not at 1.36 watts. On No. 3 amplifier where the impedance of the circuit supplying the grids is lower than in No. 2 the effect of grid modulation is still less marked on the oscillograph and the rise of third harmonic in the analysis is less rapid. This amplifier was designed for an output of .365 watts. The results show that this is obtained at the expense of the introduction of about 2% third harmonic.

The No. 4 amplifier is equivalent to the last stage of the No. 1 amplifier both using about 350 volts anode potential and 27 volts

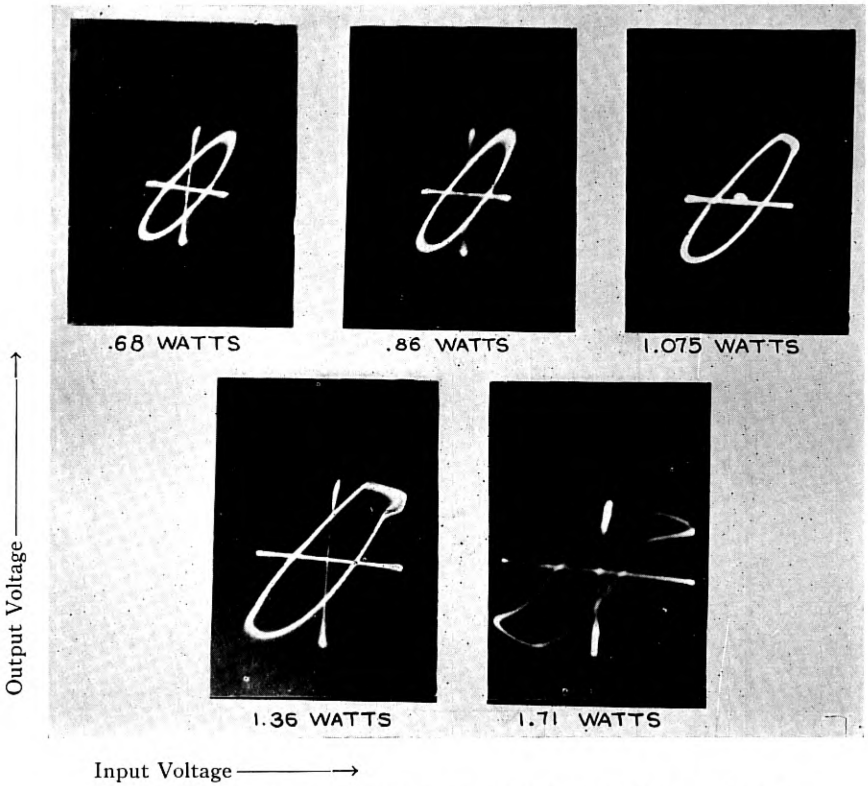


Fig. 17—Amplifier No. 4; 200 c.p.s. input; load 4,000 ohm resistance

grid bias which in the case of the No. 4 amplifier is obtained from a resistance drop in the anode circuit. The harmonic analysis curves shown in Fig. 10 indicate somewhat less second harmonic at low outputs which is probably due to the smaller number of stages. Third harmonic is approximately the same in the two cases. The oscillograph figure shows a somewhat larger output before grid modulation takes place but the difference is not great.

To check whether similar results would be obtained at other frequencies and with reactive loads oscillograms were taken with output impedances having large positive and negative phase angles at frequencies of 200 c.p.s. and 1,000 c.p.s. While the width of the ellipse obtained varied greatly as was to be expected, it was found that the points where marked irregularities in the figures occurred were at the same grid excitations as in the case of the figures taken at 1,000 c.p.s.

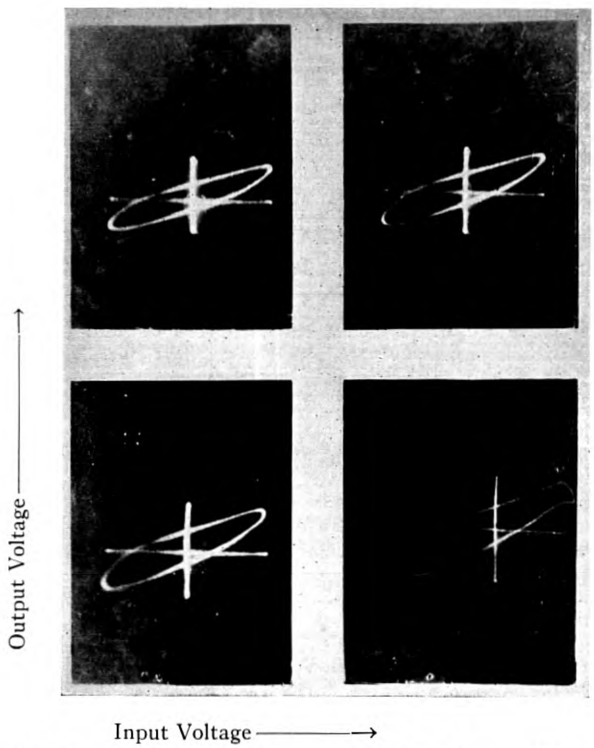


Fig. 18—Amplifier No. 4; 1000 c.p.s.; load 7,000 ohm positive reactance

with resistance load. Figures were also taken using the output current to deflect the electron stream magnetically with the same results. Typical oscillograms for these varying conditions are reproduced in Figs. 16, 17 and 18.

In conclusion, the load carrying capacity of an amplifier may be determined by either method with approximately the same results. An harmonic analysis reveals in detail the amount of each harmonic that is introduced at any load and gives useful data for fundamental studies. It requires considerable apparatus and is slow in operation. The fall of gain method while it gives approximately the same results as the other methods is not so precise since the gain falls very slowly at the overload point and does not begin to fall rapidly till the amplifier is heavily overloaded. It is therefore, difficult to pick the exact point where overloading occurs. Moreover the method affords no indication of the kind of overloading that is occurring.

Determination by the use of the cathode ray oscillograph is more rapid and in most cases more precise although in the case of push-pull amplifiers with low grid-circuit impedances the overload point is not so clearly marked as in the other cases. The shape of the curves affords valuable information as to the place in the circuit where the overloading occurs and, by comparison with previously made analyses, a good indication of the amount of harmonic introduced. It therefore forms a very valuable tool for the design engineer.

By either method the result obtained shows the load carrying capacity of an amplifier for a single frequency. The complete answer as to how much volume in speech or music a particular amplifying system will handle depends upon an analysis of the power in the speech or music such as that given in C. F. Sacia's paper on Speech Power and Energy in the October, 1925, issue of this Journal.

BIBLIOGRPAHY

1. "Physical Measurements of Audition and Their Bearing on the Theory of Hearing," Harvey, Fletcher, *Bell System Technical Journal*, Oct., 1923, Vol. II, p. 145.
2. "Speech Power and Energy," C. F. Sacia, *Bell System Technical Journal*, Oct., 1925, Vol. IV, p. 627.
3. "A Theoretical Study of the Three-Element Vacuum Tube," John R. Carson, Proceedings I. R. E., Vol. VII, No. 2.
4. "Operation of Thermionic Vacuum Tube Circuits," F. B. Llewellyn, *Bell System Technical Journal*, July, 1926, Vol. V, p. 433.
5. "Design of Non-Distorting Power Amplifiers," E. W. Kellogg, *Journal A. I. E. E.*, May, 1925, Vol. 44, p. 490.
6. "The Performance of Amplifiers," H. A. Thomas, *Journal I. E. E.* (London), Feb., 1926, Vol. 64, p. 253.
7. "Selecting an Audio-Frequency Amplifier," D. F. Whiting, *Bell Laboratories Record*, June, 1926.

Quality Control Charts[1]

By W. A. SHEWHART

IRRESPECTIVE of the care taken in defining the production procedure, the manufacturer realizes that he cannot make all units of a given kind of product identical. This is equivalent to assuming the existence of non-assignable causes of variation in quality[2] of product. Of course, random fluctuations in such factors as humidity, temperature, wear and tear of machinery and the psychological and physiologcal conditions of those individuals engaged in carrying out the manufacturing procedure may give rise to some of these apparently uncontrollable variations. Knowing this, the manufacturer contents himself with trying to produce a product which is uniform and controlled—one which does not vary from one period to another by more than an amount which may be accounted for by a system of chance or non-assignable causes producing variations independent of time.

To make clear the significance of the terms "assignable causes" and "non-assignable causes," we may make use of the following illustration. Suppose a person were to fire one hundred rounds at a target. We know what probably would happen—the individual would not hit the bull's-eye every time. Possibly some of the shots would fall within the first ring, others within the second ring, and, in general, the shots would be distributed somewhat uniformly about the center of the target. We have a more or less definite picture of some of the possible reasons why the individual would not hit the bull's-eye every time, but we probably cannot assign the reasons or causes for his missing the bull's-eye in any particular instance—the causes of missing are non-assignable. Suppose, however, that the individual tended to shoot to the right of the bull's-eye. Naturally we would conclude that there was some discoverable cause for this general tendency, i.e., we would feel that the observed effect could be assigned to some particular cause.

The reason for trying to find assignable causes is obvious—it is only through the control of such factors that we are able to improve the product without changing the whole manufacturing process. But it would be a waste of time to try to ferret out or assign some cause for a

[1] A brief description of a newly developed form of control chart for detecting lack of control of manufactured product.

[2] Quality is some function of those characteristics X, Y, Z . . ., required to define a thing. For our present purpose we shall consider that quality is a function of a single characteristic X.

fluctuation in product which is no greater than that which could have resulted from the non-assignable causes as it would be to try to find the exact manner in which each of the causes contributed to missing the bull's-eye in the analogous case of target practice just considered.

Here then is the practical commercial problem—When do the observed differences between the product for one period and that for another indicate lack of control due to assignable causes, and when, on the other hand, do the differences in quality of manufactured product observed from one period to another indicate only fortuitous, chance or random effects which we cannot reasonably hope to control without radically changing the whole manufacturing process? We shall outline a typical example of the way this question arises, outline the basis for its solution and present the results in the form of a control chart.

Typical Example

Fig. 1 shows the frequency polygon for 15,050 instruments inspected for quality X. These instruments were selected at random throughout the year from a product manufactured in quantities of approximately

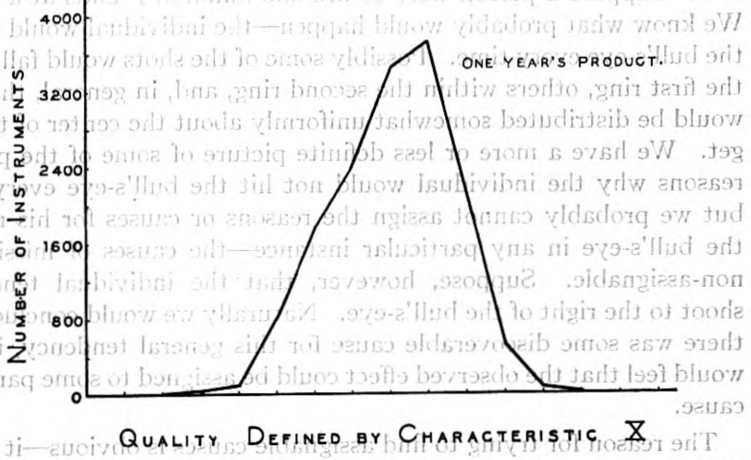


Fig. 1—Polygon showing distribution in quality for 15,050 units of product. Do these data present any evidence of lack of control?

2,000,000 per year. Is there any indication from these data that the product had not been uniform or controlled throughout the twelve month period in which the instruments had been selected?

Oftentimes we must decide from a study of a single frequency polygon of data such as given in Fig. 1, whether or not the product has been controlled during the period for which the data have been collected. In this instance, however, it was possible to group the 15,050 observations into twelve groups representing monthly samples of approximately 1250 instruments each. The data are presented in this form

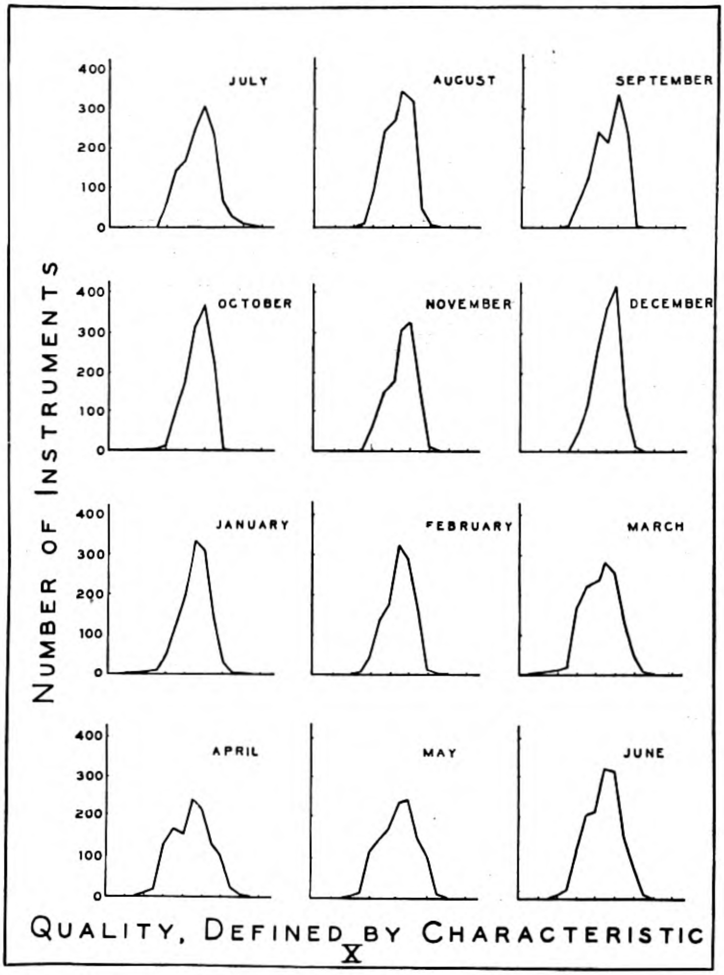


Fig. 2—Monthly polygons showing distribution in quality for samples of approximately 1250 units of product. Do these data present any evidence of lack of control?

in Fig. 2. Obviously no two polygons are the same in respect to average, dispersion and shape, but of course we would not expect them to be the same even though the product were uniform, any more than we would expect two targets to show the same distribution of shots even if the same individual had fired at both targets. In other words, non-assignable, fortuitous or chance causes introduce certain differences in the average, dispersion and shape of the observed polygons from one month to another, and we must set up some method of differentiating the effects of assignable from those of non-assignable causes.

Outline of Basis for Detecting Lack of Control

Uniform product was defined above as one for which the differences between the units or groups of units were controlled by a complex system of non-assignable chance causes producing results independent of time. Now, following a line of reasoning whose origin is attributed to Laplace, it may be shown that such a system of causes, in general, may be expected to give a unimodal distribution of product such that the probability $dy_{\lambda'}$ of the production of a unit having the quality X within the range X to $X+dX$ is independent of time, being a continuous function, f', of the quality X and certain parameters. We may represent the probability symbolically by the following equation

$$dy_{\lambda'}=f'(X,\ \lambda_1',\ \lambda_2' \ldots \lambda_{m'}')dX, \tag{1}$$

where the λ''s represent the m' parameters. Experimental evidence abounds in many fields of science to justify the adoption of Eq. 1 to represent the probability distribution of the effects of systems of chance causes. It is quite reasonable, therefore, to adopt this equation as a definition of uniform product and to use it as a basis for detecting lack of control.

Obviously, if we knew f' and the values of the m' parameters in Eq. 1, it would be comparatively easy to determine the limits within which the quality X or any estimate of a parameter derived from a sample of the product might be expected to vary because of chance causes. In practice, however, we know only the n observed values of quality obtained from inspecting a sample of as many units, and we do not know either the true functional relationship f' or any one of the m' parameters even though the product be uniform. We wish to find f' and each of the m' parameters, but, knowing that we cannot do this, we try to find some approximation f for the true function f' and some estimates $\Theta_1, \Theta_2 \ldots \Theta_m$ for the parameters $\lambda_1, \lambda_2 \ldots \lambda_m$ occurring in f. To do this we tentatively assume that the sample of n units has been drawn from a uniform product distributed in accord with the function f, and then use statistical theory to see if our assumption is justified.

Theoretically there are four fundamental steps in the procedure outlined above. They are:

1. *The Problem of Specification:* To find or specify a satisfactory form f of the distribution of the uniform product from which the sample of n pieces is assumed to have been drawn or to find the equation

$$dy_\lambda = f(X, \lambda_1, \lambda_2, \ldots \lambda_m)dX \tag{2}$$

where dy_λ is the assumed probability of a unit having a quality X within the interval X to $X+dX$.

For example we often assume the distribution to be normal so that Eq. 2 becomes

$$dy_\lambda = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-m_1)^2}{2\sigma^2}} dX, \tag{2'}$$

Here $m=2$, and λ_1 and λ_2 are respectively the arithmetic mean m_1 and the root mean square (or standard) deviation σ of X as defined by the normal curve Eq. 2′.

2. *The Problem of Estimation:* To find from the data given by the sample a suitable estimate for each of the m parameters in Eq. 2. These estimates of the parameters in terms of the data of the sample are often termed statistics. If we let Θ_i represent the chosen statistic for the parameter λ_i in Eq. 2, we may rewrite this equation as follows

$$dy_\Theta = f(X, \Theta_1, \Theta_2, \ldots \Theta_m)dX \tag{3}$$

as our theoretical approximation for the assumed true (Eq. 2) probability distribution.

An estimate of a given parameter may often be obtained in a number of ways by one or more methods.

In the above illustrative case of the normal law, we must estimate the two parameters m_1 and σ (Eq. 2′) from the n observed values of X in the sample. Now, it is well known[3] that σ may be expressed in an indefinitely large number of ways in terms of the arithmetic means of the absolute values of the integral powers of the deviations of X defined by Eq. 2′. Estimates of σ might be obtained in terms of the corresponding means calculated from the

[3] Whittaker and Robinson, Calculus of Observation, page 182.

sample. Two such estimates familiar to all are (letting Θ_2 stand in general for an estimate of σ, the second parameter of equation 2′)

$$\Theta_{21} = \sqrt{\frac{\pi}{2}}\Sigma\frac{|X-\overline{X}|}{n},$$

and

$$\Theta_{22} = \sqrt{\Sigma\frac{(X-\overline{X})^2}{n}},$$

where the summation extends over all the X's in the sample of n and $\overline{X}$ is the arithmetic mean of these values of X.

Thus for every λ occurring in Eq. 2, we may have many ways of securing an estimate from the sample. Of these ways, which one shall we choose? Obviously, as in the case of Θ_{21}, compared with Θ_{22}, one estimate may require less labor than another in its calculation. This, however, is not always the deciding factor, because one estimate may have a larger error than another. This leads us to the third problem.

3. *The Problem of Distribution:* To determine how each of the proposed estimates of a parameter might be distributed in a sequence of samples so that we may obtain some measure of its error.

In general we desire that estimate of a given parameter which has the smallest error or highest precision. Thus, in the case of Θ_{21}, it requires a sample of $1.14n$ to give as high a precision as the estimate Θ_{22} has for a sample of size n because the ratio of the error of Θ_{21} to Θ_{22} is $\sqrt{1.14}$. Hence the economic savings effected by using the better of two estimates may be very appreciable.

Furthermore the errors of the statistics are used in establishing the limits within which observed values of the statistics calculated from different samples may be expected to lie as will be illustrated below in discussing the data of Fig. 2. Naturally such errors are used in preparing the control chart Fig. 4.

Suppose now that we have taken the three steps outlined above and found the calculated or theoretical distribution in the form of Eq. 3. What assurance have we that the observed sample could have come from such a distribution? This question leads us to the fourth problem.

4. *The Problem of Fit:* To calculate the probability of fit between the observed and theoretical distributions.

Thus, if the n observed values of X are grouped into $m+1$ cells having frequencies $n_o, n_1, \ldots n_{m+1}$ and if the calculated or theoretical frequencies in these same cells as determined from Eq. 3 are $n_{o\Theta}, n_{1\Theta}, \ldots n_{m\Theta}$ where $\Sigma n_i = \Sigma n_{i\Theta} = n$, we may calculate by Pearson's method the probability P of random samples exhibiting as large or larger values of X'' than that observed in our sample where $\chi^2 = \Sigma \frac{(n_{i\Theta} - n_i)^2}{n_{i\Theta}}$. If the value of probability P thus found is small, we may conclude that it is highly improbable that the sample of n units of product came from uniform product of the form assumed. Of course, this theoretically does not settle the question as to whether the sample might have come from a uniform product other than that assumed, because, as we see, f is only an assumed form for f'. Practically, however, we seem justified in concluding that it is unlikely that the product is uniform if P is small, particularly since the choice of f is customarily made upon the basis of large samples. The application of this test is illustrated in connection with the discussion of the data in Fig. 3.

Practical Application of Theory

The application of the steps just outlined will be illustrated by an analysis of the data in Figs. 1 and 2 to show that the product had not been controlled for the period therein indicated. Carrying out steps 1 and 2 we conclude that the best theoretical equation representing the data in Fig. 1 is either[4] the Gram-Charlier series (two terms) or the Pearson curve of type IV for both of which the estimates of the parameters may be expressed in terms of the first four moments μ_1, μ_2, μ_3 and μ_4 of Fig. 3. These two distributions are shown in columns 10 and 14 respectively.[5] Pearson's test for goodness of fit (step 4) gives negligible results[6] (the probabilities of fit as measured by P on the chart are for practical purposes zero) in both instances, and this was taken as indicating that assignable causes of variation had entered the product. Further investigation of an engineering nature justified this conclusion.

We should not fail to note as suggested above, however, that a small value of fit technically indicates only that the chance is small that a random sample drawn from the theoretical universe (either the two-

[4] Equations for these curves may be found in Bowley's Elements of Statistics, pages 267 and 345 respectively.

[5] Bowley's table, page 303 in his "Elements of Statistics," was used in the calculation of the Gram-Charlier graduation.

[6] Corrections were applied to take account of the number of degrees of freedom, etc., in the calculation of goodness of fit.

OBSERVATIONS	
CHARACTERISTIC INSPECTED (a)	QUALITY X
UNITS (b)	
TO NEAREST	
SOURCE OF DATA	

INSPECTION ENGINEERING ANALYSIS SHEET

SUBJECT	
	TYPE A-INSTRUMENT
1	Quality from
2	July 1923 -
3	June 1924 incl.

CALC. BY MBC	REPORT NO.
CHK. BY MSH	DATE 1-14-26
APP'D.	
SHEETS SHEET	

$f_1(x)$ Gram-Charlier (2 terms) — $f_2(x)$ Pearson Type IV

0 Cell No.	1 Quality (a) in X (b)	2 Cell Boundaries Lower	3 Cell Boundaries Upper	4 Dev. in cells from origin at ō −5.5 X	5 Obs. Freq. y	6 yX	7 yX²	8 yX³	9 yX⁴	10 Theor. Freq. y_1	11 $y-y_1$	12 $(y-y_1)^2$	13 $\frac{(y-y_1)^2}{y_1}$	14 Theor. Freq. y_2	15 $y-y_2$	16 $(y-y_2)^2$	17 $\frac{(y-y_2)^2}{y_2}$	0 Cell No.
0	-5.5	-5.75	-5.26	0	13	0	0	0	0									0
1	-5.0	-5.25	-4.76	1	10	10	10	10	10									1
2	-4.5	-4.75	-4.26	2	8 74	16	32	64	128	86	- 12	144	1.674	103	- 29	841	8.165	2
3	-4.0	-4.25	-3.76	3	43	129	387	1161	3483									3
4	-3.5	-3.75	-3.26	4	100	400	1600	6400	25600	253	-153	23409	92.526	229	-129	16641	72.668	4
5	-3.0	-3.25	-2.76	5	815	4075	20375	101875	509375	675	140	19600	29.037	631	184	33856	53.655	5
6	-2.5	-2.75	-2.26	6	1761	10566	63396	380376	2282256	1466	295	87025	59.362	1463	298	88804	60.700	6
7	-2.0	-2.25	-1.76	7	2397	16779	117453	822171	5755197	2662	-265	70225	26.381	2708	-311	96721	35.717	7
8	-1.5	-1.75	-1.26	8	3431	27448	219584	1756672	14053376	3679	-248	61504	16.718	3716	-285	81225	21.858	8
9	-1.0	-1.25	- .76	9	3703	33327	299943	2699487	24295383	3441	262	68644	19.949	3477	226	51076	14.690	9
10	- .5	- .75	- .26	10	2165	21650	216500	2165000	21650000	2024	141	19881	9.823	997	168	28224	14.133	10
11	0	- .25	.24	11	510	5610	61710	678810	7466910	683	-173	29929	43.820	626	-116	13456	21.495	11
12	.5	.25	.74	12	77	924	11088	133056	1596672									12
13	1.0	.75	1.24	13	15 94	195	2535	32955	428415	81	13	169	2.086	100	- 6	36	.360	13
14	1.5	1.25	1.74	14	2	28	392	5488	76832									14
15																		15
	Σ				15050	121157	1015005	8783525	78143637	15050			301.376	15050			303.441	
										NO. OF CELLS = 10		P= .00000		NO. OF CELLS = 10		P= .00000		

MOMENTS ABOUT ORIGIN ō

NOTE: THE ORIGIN ō IS THE MID-VALUE BETWEEN THE BOUNDARIES OF THE CELL CHOSEN.

$${}_1\mu_1 = \frac{\Sigma yX}{\Sigma y} = \frac{121157}{15050} = 8.050299 \qquad m = \text{UNITS (b) PER CELL} = .5$$

$${}_1\mu_2 = \frac{\Sigma yX^2}{\Sigma y} = \frac{1015005}{15050} = 67.442193 \qquad \bar{X} = \bar{o} + m\,{}_1\mu_1 = -5.500000+4.025150 = -1.474850$$

$${}_1\mu_3 = \frac{\Sigma yX^3}{\Sigma y} = \frac{8783525}{15050} = 583.622924$$

$${}_1\mu_4 = \frac{\Sigma yX^4}{\Sigma y} = \frac{78143637}{15050} = 5192.268239$$

CORRECTED MOMENTS:

$$\sigma = m\,\mu_2^{1/2} = .5 \times 1.597356 = .798678$$

$$k = \frac{\mu_3}{\mu_2^{3/2}} = \frac{-1.730023}{4.075727} = -.424470$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{22.144232}{6.510387} = 3.401370$$

$$\beta_1 = k^2 = (-.424470)^2 = .180175$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} = \frac{.798678}{\sqrt{15050}} = .00651034$$

$$\sigma_\sigma = \frac{\sigma}{\sqrt{2N}} = \frac{.798678}{\sqrt{30100}} = .00460350$$

$$\sigma_k = \sqrt{\frac{6}{N}} = \sqrt{.000398671} = .0199667$$

$$\sigma_{\beta_2} = \sqrt{\frac{24}{N}} = \sqrt{.00159468} = .0399334$$

UNCORRECTED MOMENTS ABOUT ARITH. MEAN

$$\mu_1 = 0$$

$$\mu_2 = {}_1\mu_2 - {}_1\mu_1^2 = 67.442193-64.807314 = 2.634879$$

$$\mu_3 = {}_1\mu_3 - 3\,{}_1\mu_1\,{}_1\mu_2 + 2\,{}_1\mu_1^3 = 583.622924-1628.789457+1043.436510 = -1.730023$$

$$\mu_4 = {}_1\mu_4 - 4\,{}_1\mu_1\,{}_1\mu_3 + 6\,{}_1\mu_1^2\,{}_1\mu_2 - 3\,{}_1\mu_1^4 = 5192.268239-18793.356164+26224.484274-12599.963844=23.432505$$

$$3\sigma_{\bar{X}} = .0195310$$

$$3\sigma_\sigma = .0138105$$

$$3\sigma_k = .0599001$$

$$3\sigma_{\beta_2} = .119800$$

CORRECTED MOMENTS ABOUT ARITH. MEAN (SHEPPARD'S CORRECTIONS)

$$\mu_2(\text{cor.}) = \mu_2 - \frac{h^2}{12} = 2.634879-.08333 = 2.551546$$

$$\mu_4(\text{cor.}) = \mu_4 - \frac{1}{2}h^2\mu_2 + \frac{7}{240}h^4 = 23.432505-1.317440+.029167 = 22.144232$$

$$h = 1$$

Fig. 3

term Gram-Charlier series or Pearson IV type in this case) would give as large or larger value of χ^2 than that observed. Therefore the basis for the conclusion at the end of the previous paragraph is that we have faith[7] that the customary method of taking theoretical steps 1 and 2 gives a close approximation to the true distribution of the product when it is uniform or controlled.

Turning to a study of the data grouped into monthly distributions (Fig. 2), we find additional evidence of lack of control. Naturally the monthly observed values of the four statistics, average $\overline{X}$, standard deviation σ, skewness $k=\sqrt{\beta_1}$, and kurtosis β_2 should lie within well-defined limits established by sampling theory (step 3) and shown in Fig. 4, if the product had been controlled. Furthermore, the observed values of percentage defective p (percentage of instruments having quality less than some value X) from month to month also should fall within well-defined limits. Using the grand average[8] of a statistic as the basis for establishing limits, the first five sections of the control chart in Fig. 4 were constructed. The dotted lines calculated upon the basis of a uniform sample of 1250 indicate the limits within which the different statistics should lie, if the product had been controlled. The chart shows that observed values of these statistics often fall outside their respective limits indicating, subject to limitations imposed by the method of calculation, lack of control of product.

We may go still further and, without carrying out the analysis of Fig. 3, make use of Pearson's test of goodness of fit to calculate the probability that the first two months' samples could have been drawn from the same universe (the same uniform product), then that the third month's sample could have come from the same universe as the combined samples for the first and second months, etc.[9] Obviously the values of χ^2 used as a basis for this calculation of the goodness of fit

[7] Such faith may be based upon the a priori conception that an observed difference in two values of X is the resultant effect of a large number of causes (following in the steps of Laplace, Charlier, Edgeworth, Gram, Thiele and others) and upon the experience that observed homogeneous distributions always have been fitted by some one of the well-known forms of probability curves (following in the steps of Pearson and others).

[8] Some objection may be raised to the use of the observed average as a basis for establishing the limits of a given statistic, because this observed average almost certainly would not be the true value even though the product had been uniform. In the present case, however, we are probably justified in using the observed average because previous experience based upon thousands of observations has given approximately the same values for these quantities. Rigorously, of course, we should find the standard deviations of monthly differences from the grand average and set up limits on this basis. Wherever necessary this method is followed and in fact has been carried out for the case in hand where it gives results similar to those indicated in Fig. 4.

[9] Pearson, K.P., Biometrika, vol. viii, 1911, p. 250 and vol. x, 1914, p. 85.
Rhodes, E.C., Biometrika, vol. xvi, 1924, p. 239.

should fall within well-defined limits such as indicated on the chart. Reference to the χ^2-part of the control chart, Fig. 4, shows that this test gives more conclusive evidence than any other for deciding that the product had not been controlled. As previously noted, further investigation revealed the assignable causes of lack of control. This is a

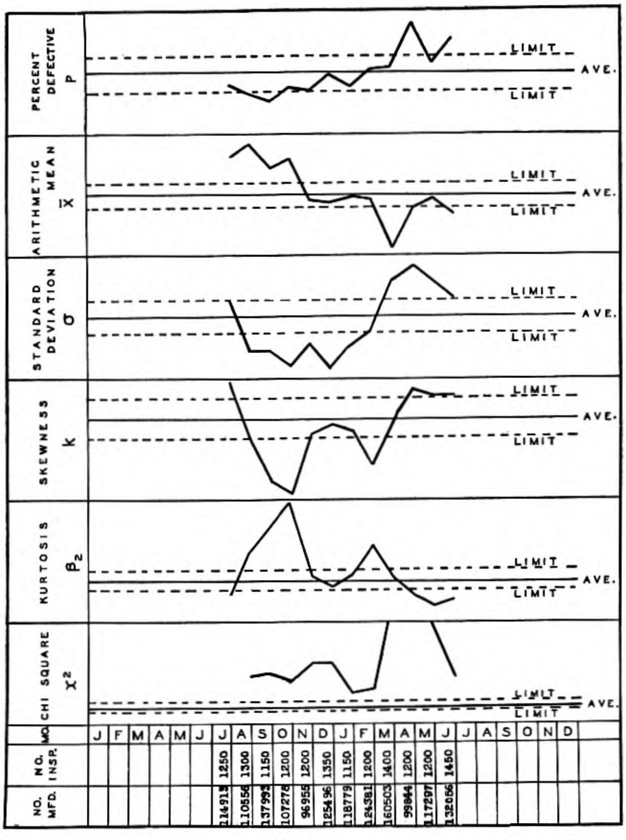


Fig. 4

common experience under such circumstances. Furthermore, it is of interest to note that the preparation of such a chart requires but a small amount of labor on the part of a computer.

Discussion and Conclusion

This paper shows how statistical methods may be used to detect lack of control of product. It describes a recently developed form of manufacturing control chart which helps in the use of inspection and production data by applying some of the modern tools of the statistician. The chart tells the manufacturer at a glance whether or not the product has been controlled. Evidence of lack of control calls for immediate attention, but there need be no time lost in looking for causes of variation in product when these variations are not large enough to indicate lack of control.

There is an obvious advantage in using all parts of the chart wherever possible, because, as the illustration shows, one part may reveal trouble even though some other parts do not. However, when the inspection is made on the basis of attributes, the data will be available for the first or percentage defective part of the chart only.

Applications of Poisson's Probability Summation

By FRANCES THORNDIKE

SYNOPSIS: The applicability of Poisson's exponential summation to a variety of actual data is illustrated by thirty-two examples of actual frequency-distributions to which the Poisson distribution is a fairly good approximation. The comparison of actual and theoretical distributions is made graphically, using as a background new probability curves showing Poisson's exponential summation with a logarithmic scale for the average. To suggest possible explanations of the observed deviations from the theoretical Poisson distribution consideration is given to the effect on the theoretical distribution of certain modifications in the underlying assumptions, corresponding to conditions under which much actual data must be obtained.

IN an earlier number of THE BELL SYSTEM TECHNICAL JOURNAL there were published two sets of curves showing Poisson's exponential summation.[1] These charts, which are shown on a reduced scale in Figs. 1 and 2, give the relation between a, the average number of occurrences of an event in a large group of trials, the number of trials being very great compared with the average a, and the probability P that the actual number of occurrences in any such group of trials will equal or exceed any given number c. The purpose of this paper is to facilitate the use of these curves by making clear the characteristics of the Poisson summation, especially the assumptions on which it is based, and the precautions which must be observed in applying it, these points being illustrated by a number of actual frequency-distributions for which the Poisson distribution furnishes a fairly good working approximation.

POISSON'S EXPONENTIAL SUMMATION

Three assumptions underlie the mathematical treatment of Poisson's exponential summation

$$P=1-\left[1+\frac{a}{1!}+\frac{a^2}{2!}+\frac{a^3}{3!}+\ \dots\ +\frac{a^{c-1}}{(c-1)!}\right]e^{-a}$$

and its application to practical problems. The first is that the quantity measured is the number of occurrences of a particular event which always definitely happens or fails to happen, so that the actual number of occurrences c is either zero or a positive integer. The second assumption is that we may imagine the group of trials con-

[1] Figs. 1 and 2 of "Probability Curves Showing Poisson's Exponential Summation," by G. A. Campbell, *Bell System Technical Journal*, Vol. 2, No. 1, pp. 95–113, January, 1923.

stituting the sample in question to be repeated an infinite number of times, independently and uniformly, with an average number of occurrences per sample equal to a, so that we may speak of a as the average number of occurrences for the sample in question. The

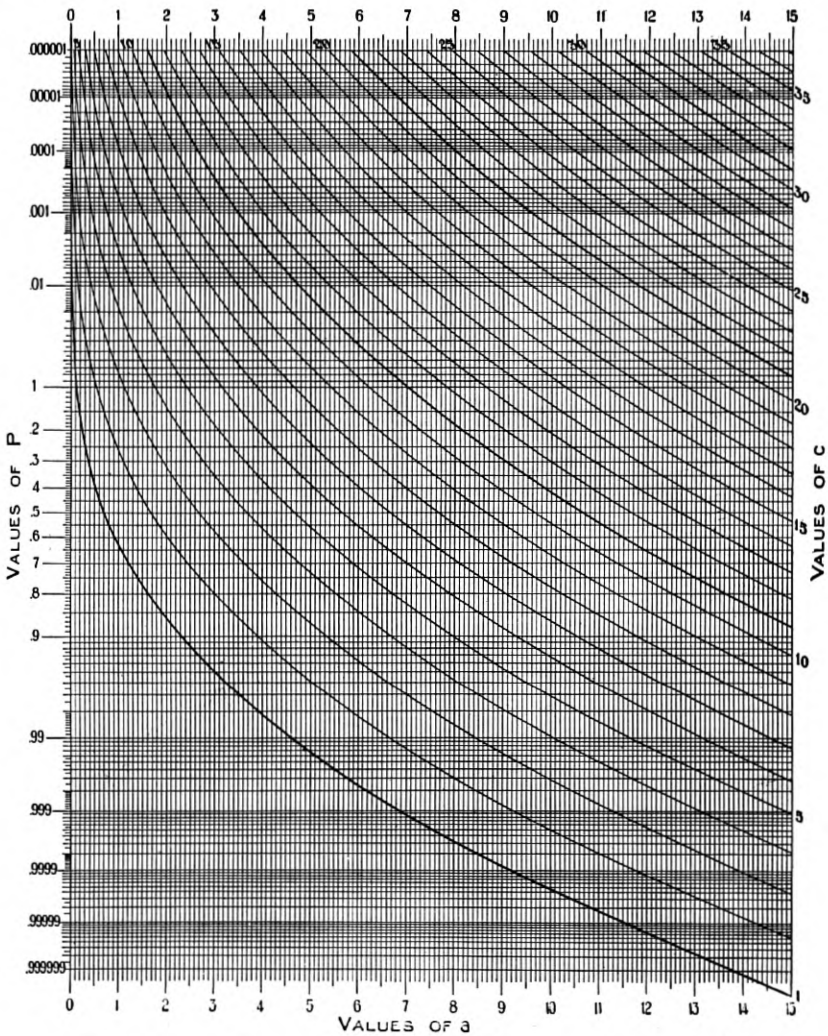


Fig. 1—Probability curves showing Poisson's exponential summation

$$P=1-\left[1+\frac{a}{1!}+\frac{a^2}{2!}+\ldots+\frac{a^{c-1}}{(c-1)!}\right]e^{-a}$$

for the probability P that an event occur at least c times in a large group of trials for which the average number of occurrences is a. A scale proportional to the normal probability integral is used for P, a linear scale for a

third assumption is that, while the sample has a finite average number of occurrences, it consists of an infinite number of independent, uniform trials, so that the possible number of occurrences in a sample is infinite, and the probability that the event occur in a single trial

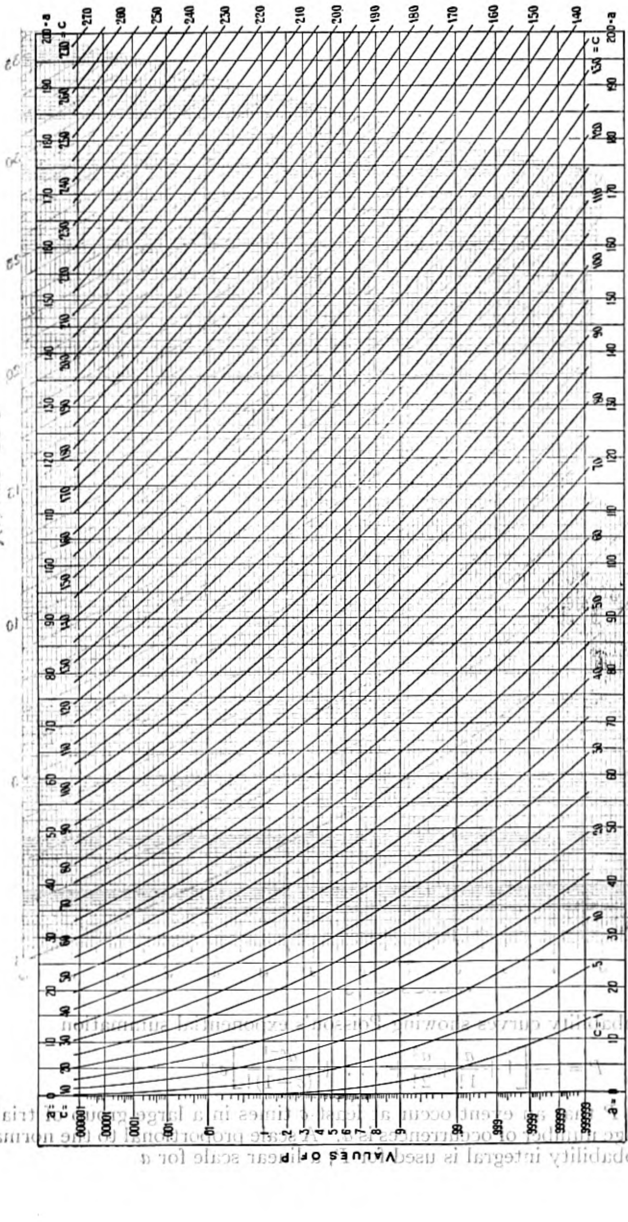


Fig. 2—Probability curves showing Poisson's exponential summation

$$P = 1 - \left[1 + \frac{a}{1!} + \frac{a^2}{2!} + \ldots + \frac{a^{c-1}}{(c-1)!}\right] e^{-a}$$

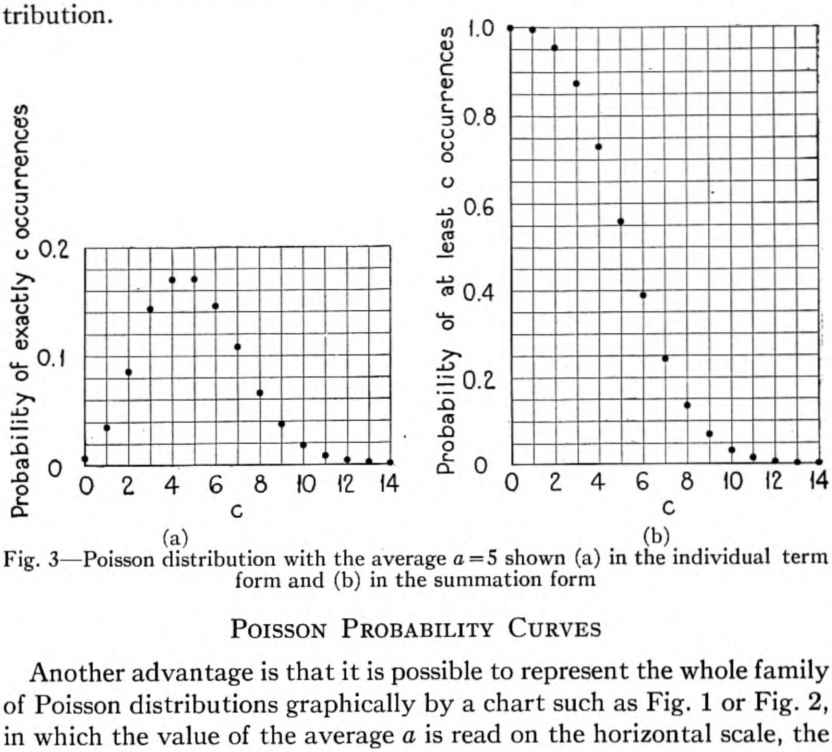
for the probability P that an event occur at least c times in a large group of trials for which the average number of occurrences is a. A scale proportional to the normal probability integral is used for P, a linear scale for a

is infinitely small. The term "uniform" applies, of course, not to the results of the trials (or samples) but to the essential conditions under which they are obtained, and "independent" is used with the meaning that the result of one trial (or sample) does not affect the occurrence of the event in any other trial (or sample). The first and third assumptions, translated into exact mathematical language, define a particular kind of probability function, which can be derived by taking the limit, as n becomes infinite and pn remains finite, of the point binomial $(p+q)^n$ for the probability of any number of occurrences of a given event in a group of n independent, uniform trials, when the probability that the event occur in a single trial is p. The second assumption is required in order that we may pass from the abstract idea of a probability function to the concrete idea of a frequency-distribution.

Throughout this discussion the summation form of the frequency-distribution, giving the probability of at least c occurrences, is used rather than the individual term form, giving the probability of exactly c occurrences. One reason for the use of the summation form is its more direct applicability to many practical problems in which the chance of exceeding a certain limit, rather than the chance of obtaining any one particular value, is of practical importance. Secondly, as Fig. 3a shows, the individual term form gives in general two possible values of c for any pair of values of a and P, whereas the summation form is single-valued and introduces no such ambiguity.

Fig. 3 also calls attention to some of the outstanding characteristics of the Poisson distribution, its discontinuity and skewness, in particular. That the Poisson distribution must be a series of discrete points and not a continuous curve is a direct result of the assumption that c represents a number of occurrences. That the distribution is skew follows from the fact that the possible number of occurrences is much larger, in fact infinitely larger, than the average number of occurrences. This skewness is quite marked even in the Poisson distribution with $a=5$, which is shown in Fig. 3, and it becomes more pronounced as a is decreased toward zero. If, for example, the average number of occurrences in a million trials is one, in any particular group of a million trials it is equally likely that there will be no occurrence of the event or one occurrence, and it is almost 1.4 times as likely that there will be no occurrence as that there will be two or more occurrences, though zero and two are equally removed from the average. A third important characteristic of the Poisson exponential, which is not brought out by this figure, is its extreme simplicity. The distribution is entirely determined by the value given to a single parameter, the average a; its standard deviation is $\sqrt{a}$, its skewness is $1/\sqrt{a}$, and its kurtosis is $3+1/a$.[2]

One consequence of this simplicity is that there is no difficulty in deciding on a definition of the *corresponding Poisson distribution* with which any other distribution should be compared. It is naturally the Poisson distribution having the same average as the given distribution.



(a) (b)

Fig. 3—Poisson distribution with the average $a=5$ shown (a) in the individual term form and (b) in the summation form

Poisson Probability Curves

Another advantage is that it is possible to represent the whole family of Poisson distributions graphically by a chart such as Fig. 1 or Fig. 2, in which the value of the average a is read on the horizontal scale, the value of the probability P on the vertical scale, and the number of occurrences c on the individual curves of the set. Any two of these three variables may then be taken as the independent variables and the values assigned to them will determine the value of the third variable, which can be read off at once. The only ambiguity occurs

[2] The standard deviation (σ), skewness (k), and kurtosis (β_2) of any distribution are defined as

$$\sqrt{\frac{\Sigma(x_i-a)^2}{N}},\quad \frac{\Sigma(x_i-a)^3}{\sigma^3},\ \text{and}\ \frac{\Sigma(x_i-a)^4}{\sigma^4},$$

respectively, N being the number of samples in the series, and x_i the actual number of occurrences in the ith sample. For any point binomial

$$\sigma=\sqrt{npq},\quad k=\frac{q-p}{\sqrt{npq}},\quad \beta_2=3+\frac{1-6pq}{npq}.$$

when a and P are the independent variables. The point determined by their values will, in general, fall between two of the c curves and the interpretation of P must be known to determine which of the two values of c should be taken. The desired value of c is read from the lower curve if P means *a probability of P or more*, from the upper curve if P means *a probability of not more than P*.

These charts may then be used conveniently in place of unwieldy double-entry tables to obtain theoretical values needed either for comparison with experimental data or to take the place of experimental data. Examples of such uses of the Poisson exponential are discussed in detail by Karl Pearson,[3] W. A. Shewhart,[4] and E. C. Molina.[5] The use of these curves in the study of telephone trunking, letting a represent the average number of simultaneous calls from a large group of subscribers, $c-1$ the number of trunks provided for them, and P the probability that all the trunks will be in use when a subscriber attempts to make a call, is suggested by Mr. Molina's paper. Other possible applications might be found in connection with the control of errors in service, defects in a manufactured article, the stock on hand of staple articles such as ink, shoe-polish, or spark plugs, or the number of copies of reference books in a library serving a large number of people. Still others may be suggested by Table I, which is a summary of the actual data now brought together for the first time for comparison with the theory.

The comparison of any actual distribution with the corresponding Poisson distribution may easily be made graphically, using these curves as a background. In fact the charts will often be found useful as coordinate paper on which to plot any frequency-distribution, theoretical or observed, provided the values of the variate are inherently limited to the positive integers and zero.

When the curves are used in this way the corresponding Poisson distribution is represented by the points in which the vertical line for the observed value of a cuts the c curves, or for convenience simply by the vertical line itself. The other distribution may then be plotted with c and P as the independent variables, and the horizontal deviations of these points from the vertical line serve as a measure of the discrepancy between the two distributions.[6] If the comparison is to be made with an observed frequency-distribution the values used

[3] Introduction to "Tables of the Incomplete Gamma Function," London, 1922.

[4] "Some Applications of Statistical Methods to the Analysis of Physical and Engineering Data," *Bell System Technical Journal*, Vol. 3, No. 1, pp. 43–87, January, 1924.

[5] "The Theory of Probabilities Applied to Telephone Trunking Problems," *Bell System Technical Journal*, Vol. 1, No. 2, pp. 69–81, November, 1922.

[6] The distributions might be plotted in other ways, e.g., letting P or c be the dependent variable, but the method used here is the simplest.

TABLE I

N = number of samples
aN = total number of occurrences
a = average number of occurrences per sample

	Series	N	aN	a
a 1	Alpha particles	2608	10097	3.87
a 2	Alpha particles	1304	10094	7.74
a 3	Deaths of aged	1096	903	0.82
a 4	Deaths of aged	1096	2364	2.16
a 5	Telephone lines in use	>1000	>4315	4.32
a 6	Bacilli	1000	1927	1.93
b 1	Yeast cells	400	720	1.80
b 2	Yeast cells	400	1872	4.68
b 3	Lost articles	423	439	1.04
b 4	Number *12*	500	421	0.84
b 5	Fires	364	9487	26.1
b 6	Incorrect reports	506	138	0.27
b 7	Cutoffs	506	1057	2.09
b 8	Double connections	506	1760	3.48
b 9	Calls for wrong number	506	2520	4.98
c 1	Deaths from kick of horse	200	122	0.61
c 2	Number *12*	250	251	1.00
c 3	Calls from group of two coin-box telephones	145	172	1.19
c 4	Calls from group of four coin-box telephones	140	384	2.74
c 5	Calls from group of two coin-box telephones	141	212	1.50
c 6	Calls from group of six coin-box telephones	138	468	3.39
c 7	Cutoffs	267	557	2.09
c 8	Double connections	267	906	3.39
c 9	Calls for wrong number	267	1351	5.06
c10	Connections to wrong number	267	2334	8.74
c11	Party lines	300	1981	6.60
c12	"Lost and found" advertisements	209	7051	33.7
d 1	Number *12*	100	421	4.21
d 2	Number *12*	50	421	8.42
d 3	Comets	100	258	2.58
d 4	Particles in emulsion	50	46	0.92
d 5	Particles in emulsion	50	106	2.12

for the probability P are the values of the observed relative frequency F, which are calculated as indicated in Table II, and the observed distribution is represented by an irregular series of dots, as in Fig. 4.

A third set of curves, Fig. 5, supplementary to Figs. 1 and 2, has now been drawn using a logarithmic scale for a. This chart shows the individual c curves up as far as $a=30$ and it shows more clearly than does Fig. 1 the range $0.1 \leqq a \leqq 2$. It may also be used as a background in the same way as Figs. 1 and 2, with the additional advantage of making the distances of the plotted points from the vertical line proportional to the percentage deviations rather than proportional to the absolute values of the deviations, so that the fit of a distribution having a small average can be compared directly by eye with that of a distribution having a large average, since it is more often the relative than the absolute value of the deviation which is significant.

Practical Applications

In applying the Poisson summation to any concrete problem, or in comparing any observed distribution with the corresponding Poisson distribution, it is necessary to bear in mind several practical conditions which must work against any perfect agreement between the

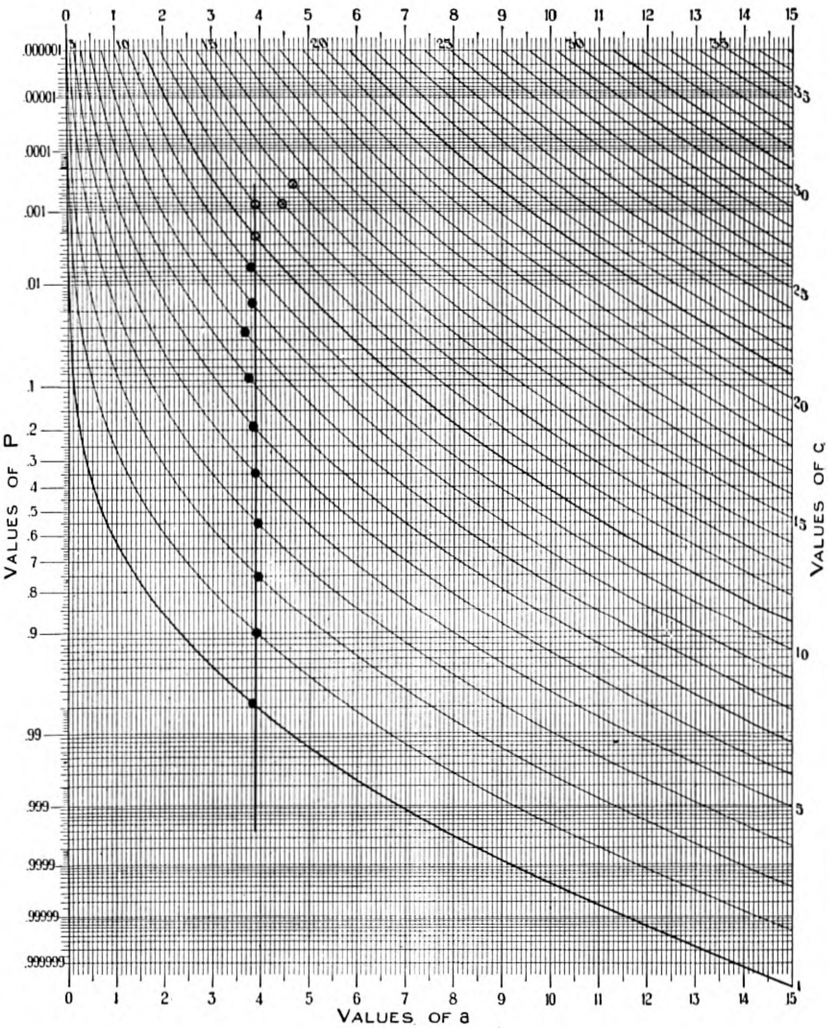


Fig. 4—Comparison of an observed distribution of the number of α particles emitted with the corresponding Poisson distribution, showing the method of using Fig. 1 or Fig. 2 as a background for plotting actual distributions. The Poisson distribution is shown by a vertical line, the observed distribution by dots

observed distribution and the corresponding Poisson distribution. In the first place, the sample considered will necessarily consist of a finite number of trials instead of an infinite number as assumed in the mathematical theory, and the trials may not be completely independent or entirely uniform. Secondly, even if the individual sample possessed the ideal characteristics assumed in the mathematical formulation, the actual series of samples must be finite and the samples may be interdependent and far from uniform. The size of the samples relating to the economic, geographic, and time divisions ordinarily used in statistical work generally varies considerably. The effect of modifying the original mathematical assumptions to correspond with some of these actual conditions is illustrated by Figs. 6–8, which show various theoretical frequency-distributions plotted on Fig. 1 or Fig. 2 for comparison with the corresponding Poisson distributions.

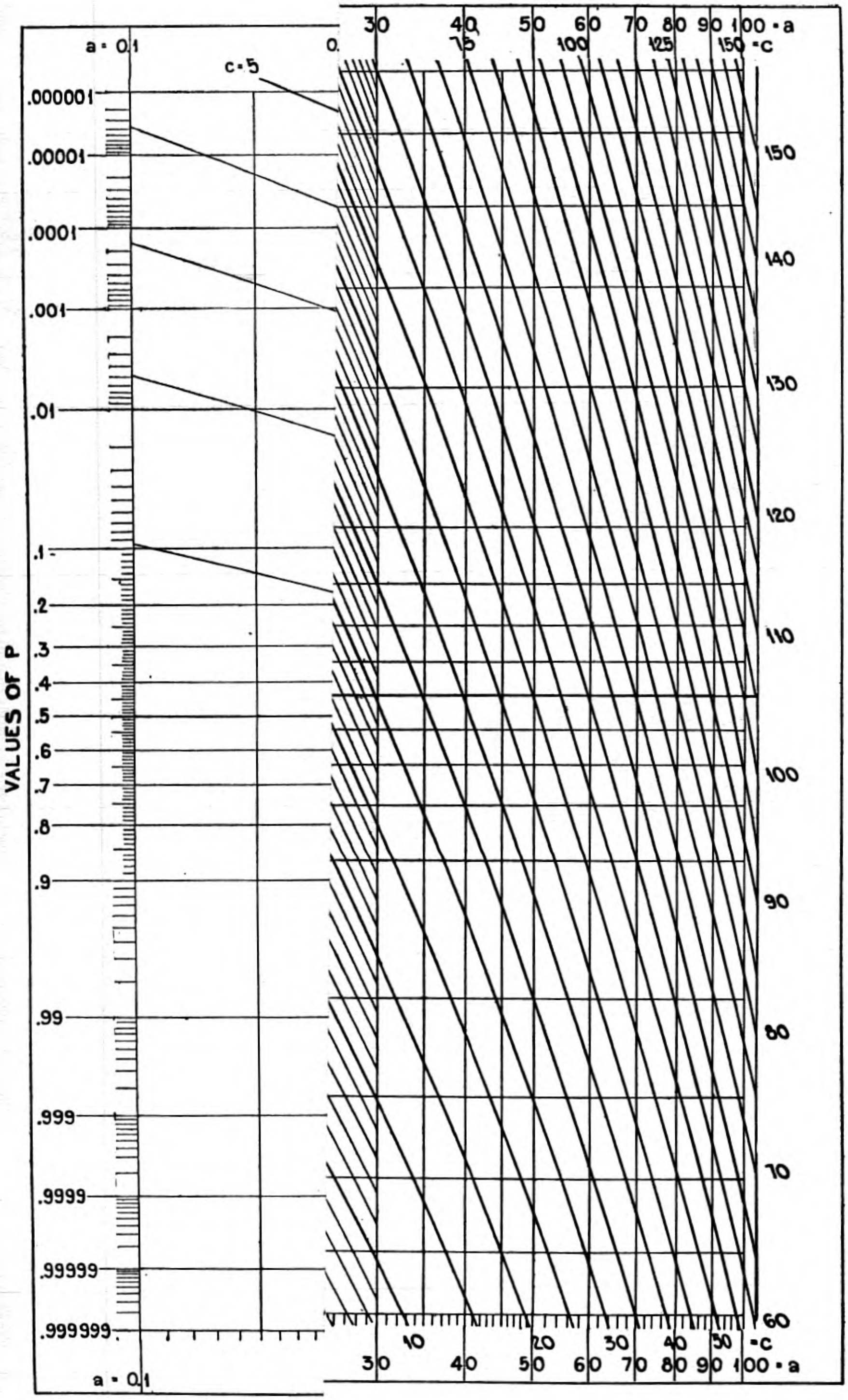
The finiteness of the number of trials n not only makes impossible the occurrence of values of c greater than the value of n, but also tends to produce a general trend away from the Poisson distribution. This is illustrated by the four typical finite binomial distributions shown in Fig. 6, which have a definite curve and slope toward the left which becomes more pronounced as n is decreased.[7] Interdependence of the trials constituting a sample will also tend to give the resulting distribution a slant, to the right if the correlation is positive, to the left if the correlation is negative.[8] Thirdly, even though the trials are independent, if they are not uniform, there will be a tendency for the distribution to slant to the left.

The requirement that N, the number of samples in the actual series, be finite introduces a somewhat different kind of deviation from the theoretical Poisson distribution. The observed relative frequency F, which is compared with the theoretical probability P, is an integral multiple of $1/N$, so that, since N is finite, the points representing the observed distribution (except those at $P=0$ and $P=1$, for which the ordinates are plus and minus infinity, and which, therefore, never appear on the graph) are all in the finite range between the two horizontal lines $P=1/N$ and $P=1-1/N$. Not only is the occurrence of points outside this range impossible, but the points near its extremes, being determined by a comparatively small number of samples, are of less significance than those near the center.

To call attention to these facts all observed distributions shown here have been represented, as in Fig. 4, with the vertical line rep-

[7] A more detailed discussion of the effect of finite sampling will be found in the paper by G. A. Campbell previously referred to.

[8] See "Explanation of Deviations from Poisson's Law in Practice," by "Student," *Biometrika*, Vol. 12, pp. 211–215, 1919.



tional to the

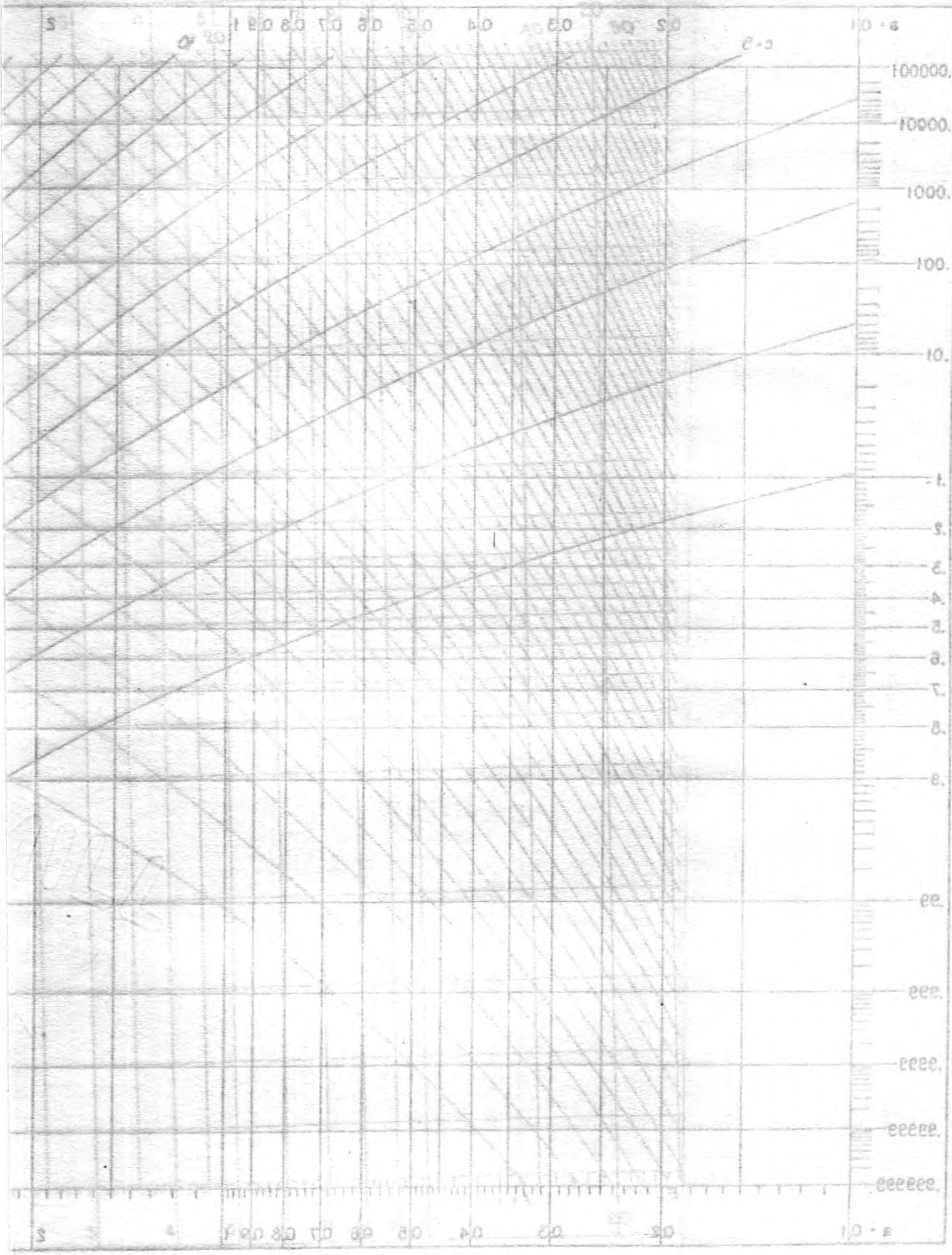


Fig. 6—Probability cur[ves]

$$P = 1 - e^{-a}\left[1 + \frac{a}{1!} + \frac{a^2}{2!} + \cdots + \frac{a^{c-1}}{(c-1)!}\right]$$

for the probability P that an event occur at least c times in a large gr[oup] ... normal probability in[tegral] resenting the corresponding Poisson distribution terminated at $P=1/N$ and $P=1-1/N$, and with the observed points in the range $P=10/N$ to $P=1-10/N$ shown as solid black dots and the points outside this range shown as circles with white centers. This secondary division is quite arbitrary, for the increase in reliability of the points as the center of the range is approached is gradual. There will, of course, be irregularities due to sampling even in the center as long as the number of samples is finite.

Non-uniformity of the samples of the series may introduce a definite trend away from the Poisson distribution, a slant to the right such as is shown in Figs. 7 and 8. Such trends result when the value of a varies from sample to sample of the series. Fig. 7 shows three theoretical distributions of this sort, each having the same average $a=75$. Series (a) is made up of two equal sub-series having $a=50$ and $a=100$, respectively, (b) of two unequal sub-series, in the ratio of 3:1, having $a=60$ and $a=120$, respectively, and (c) of three equal sub-series having $a=15$, $a=60$, and $a=150$, respectively.[9] Fig. 8 shows the effect on the distribution of letting a vary continuously and uniformly between the limits 5 and 15, the compound series (b) made up of two equal sub-series with averages 5 and 15 being also shown for comparison.[10] Since in practical time series a usually increases or decreases with the time, this kind of distribution may be expected to occur frequently. It should be noted that in all these cases it is immaterial whether a changes because of a change in the number of trials in the sample, or because of a change in the probability of the event's happening at a single trial, or because of both; if a is constant throughout the series a Poisson distribution will be obtained, and if a varies the tendency to slope to the right will be introduced. Various devices may be employed to keep the average constant in an actual series, some of which will be illustrated by the examples given below.

In selecting the following examples of the Poisson summation only two general rules were followed: that there must be some reason to

[9] In a compound distribution

$$P=\sum \frac{N_i}{N} P_i$$

where N_i is the number of samples with the average a_i, and $P_i=P(c, a_i)$.

[10] If a varies uniformly and continuously from a_1 to a_2

$$P=\int_{a_1}^{a_2} \frac{P(c, a)}{a_2-a_1} da$$

$$=1-\frac{1}{a_2-a_1} \sum_{i=0}^{c} [P(i, a_2)-P(i, a_1)].$$

suppose the possible number of occurrences n to be at least thirty times the average a and at least 25, and that N, the number of samples in the series, must be at least 50. This last requirement excludes from our list a number of series which have previously been presented

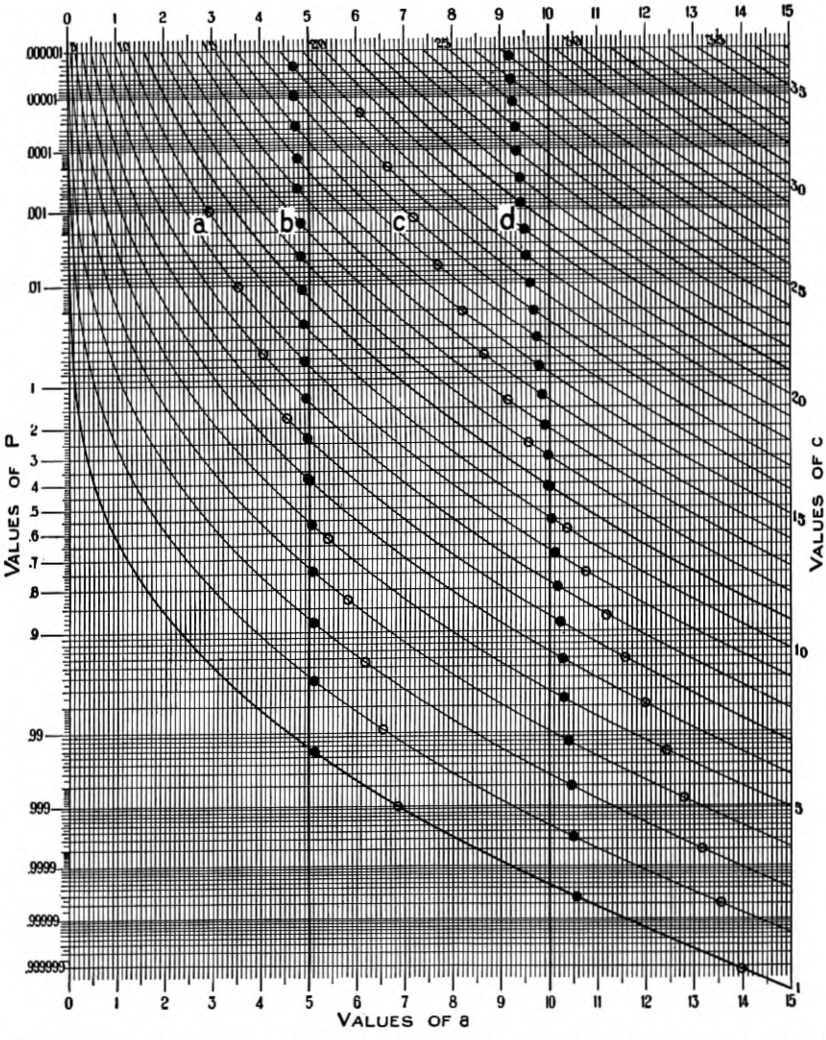


Fig. 6—Typical finite binomial distributions for the probability that an event occur at least c times in a group of n trials for which the average number of occurrences is $a = np$

(a) $a = 5$, $n = 10$
(b) $a = 5$, $n = 100$
(c) $a = 10$, $n = 20$
(d) $a = 10$, $n = 100$

as examples of the Poisson exponential, in particular those of Mortara [11] and all but one of those given by Bortkewitsch.[12]

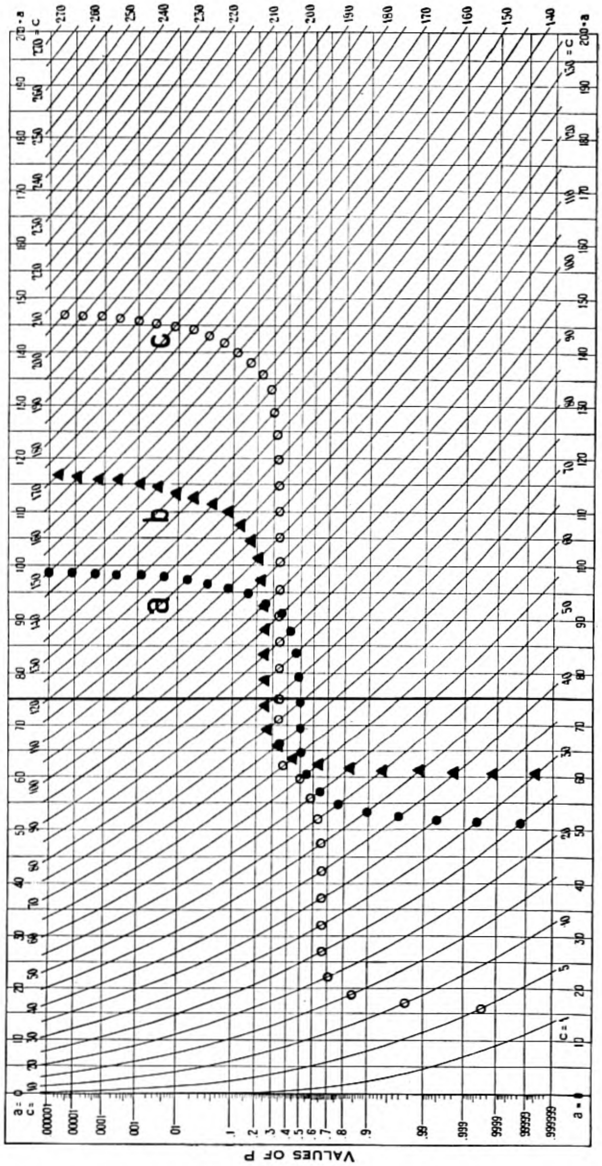


Fig. 7—Theoretical distributions for series in which
(a) $a=50$ for one half of the samples and $a=100$ for the other half
(b) $a=60$ for three quarters of the samples and $a=120$ for the other quarter
(c) $a=15$ for one third of the samples, $a=60$ for one third, and $a=150$ for the other third

[11] "Sulle Variazione di Frequenza di Alcuni Fenomeni Demografici Rari," by Giorgio Mortara, *Annali di Statistica*, Series V, Vol. 4, pp. 5–61, 1912.

[12] "Das Gesetz der kleinen Zahlen," by L. von Bortkewitsch, Leipzig, 1898.

Each of the thirty-two actual distributions shown in Fig. 9 has been plotted using Fig. 5 as the background, so that the percentage deviations in all distributions may be compared directly by inspection without regard to the magnitude of the average. The examples are divided into four groups according to the number of samples in the series, and are arranged in each group roughly in order of decreasing agreement of the observed with the theoretical distributions. A

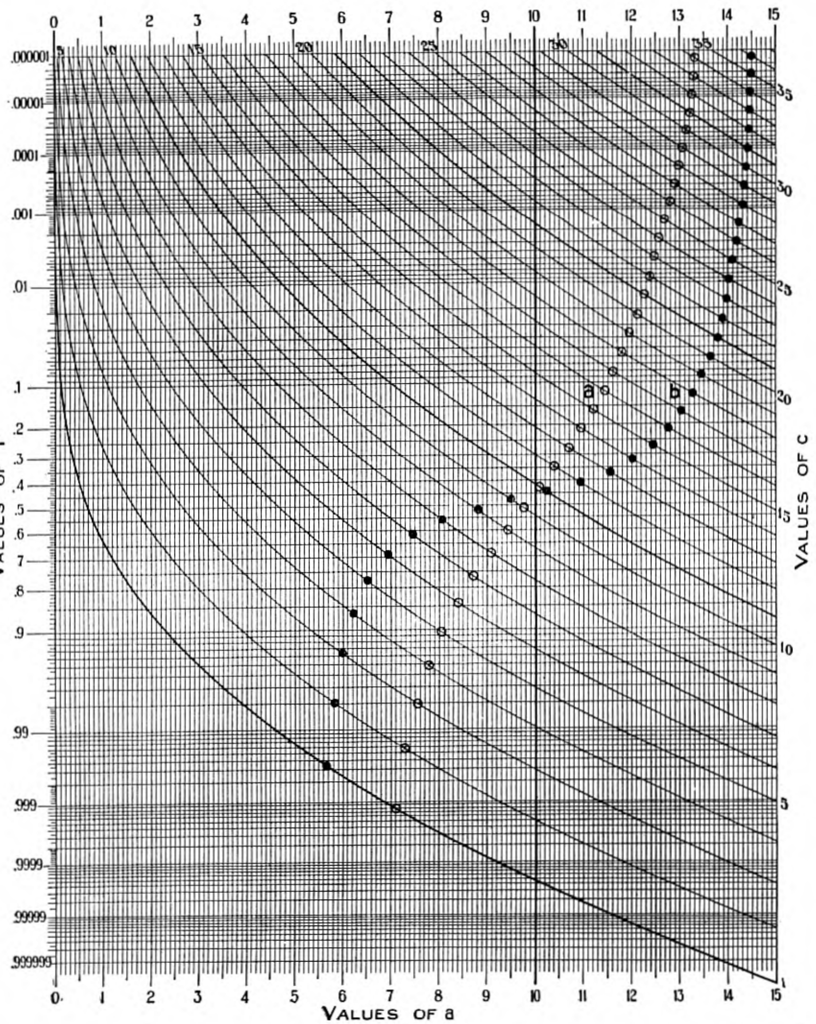


Fig. 8—Theoretical distribution for a series in which the average a varies continuously and uniformly from 5 to 15

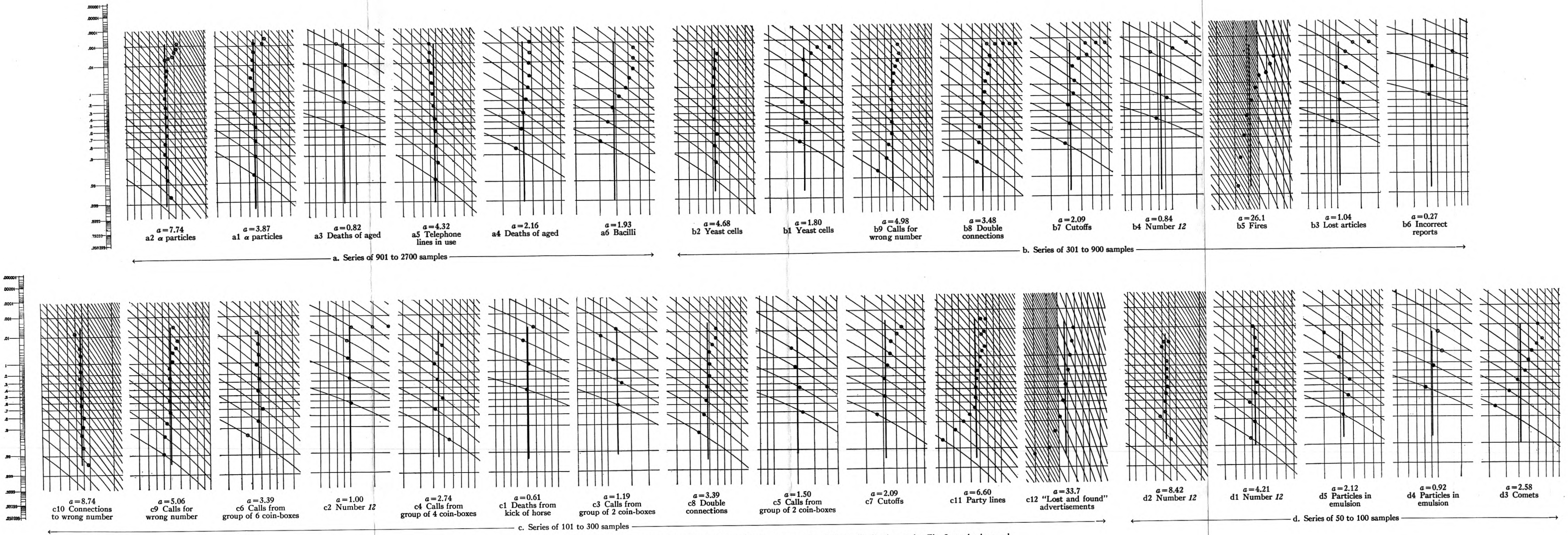


Fig. 9—Comparison of observed distributions with the corresponding Poisson distributions, using Fig. 5 as a background

summary of the data used is given in Table I and the observed distributions are given in full in Table II.

The distributions shown in the first group are taken from the work of Rutherford and Geiger, Whitaker, Holm, and Greenwood and White. Rutherford and Geiger observed the collision with a small screen of an α particle emitted from a small bar of polonium placed at a short distance from the screen. The number of such collisions in each of 2608 eighth-minute intervals was recorded, the distance between bar and screen being gradually decreased so as to compensate for the decay of the radioactive substance. From this record two frequency-distributions were calculated, that of the number of particles striking the screen in an eighth-minute interval, and in a quarter-minute interval.[13] These are distributions (a1) and (a2), respectively. Distributions (a3) and (a4) are based on a count of the number of death notices in the London *Times* on each day for three consecutive years.[14] The distribution of deaths of men over 85 years of age (a3) and that of deaths of women over 80 (a4) are shown here. The next (a5) is a frequency-distribution of the number of telephone lines simultaneously in use, from measurements on a group of 100 subscribers.[15] The last distribution of this group (a6) was obtained from a count of the number of bacilli in each of 1,000 phagocytes, or white blood cells, in the same solution and as far as possible under the same conditions, and is typical of a large number of distributions of the number of tubercle bacilli ingested per cell.[16]

The first two examples in the second group are due to "Student" and the remaining seven are new. Distributions (b1) and (b2) show the results obtained from two different solutions of yeast cells by counting the number of cells per square of a haemacytometer slide on which the solution had been spread as uniformly as possible after it had been thoroughly shaken to break up any clumps of cells.[17] The next example (b3) was obtained from the records of the "lost and found" office of the Telephone and Telegraph Building, 195 Broadway, New York City. The number of lost articles found in the building

[13] "The Probability Variations in the Distribution of α Particles," by Ernest Rutherford and Hans Geiger, *Phil. Mag.*, Vol. 20, pp. 698–707, October, 1910.

[14] "On the Poisson Law of Small Numbers," by Lucy Whitaker, *Biometrika*, Vol. 10, pp. 36–71, 1914. Six other similar distributions are given.

[15] "Calculation of Blocking Factors of Automatic Exchanges," by Ragnar Holm, *P. O. E. E. J.*, Vol. 15, pp. 22–38, April, 1922.

[16] "A Biometric Study of Phagocytosis with Special Reference to the 'Opsonic Index'," by M. Greenwood and J. D. C. White, *Biometrika*, Vol. 6, pp. 376–401, 1908–1909. Fourteen other distributions are given.

[17] "On the Error of Counting with a Haemacytometer," by "Student," *Biometrika*, Vol. 5, pp. 351–360, 1906–1907. Two other distributions are given.

and turned in to the office on each day except Sundays and holidays was recorded and tabulated for the period from November 1, 1923 to September 30, 1925, inclusive, excluding June, July, and August of each year, when there might be considerable variations in the population of the building. Distribution (b4) shows the result of a count of the number of times that the number *12* appeared as the last two digits of a ten-place logarithm in a sample consisting of a column of 100 logarithms in Duffield's table,[18] and (b5) shows the number of fires per day in New York City in 1924, as reported daily in *The New York Times*, the figures for July 4 and for Election Day being discarded for obvious reasons. The last four examples in this group were taken from telephone company records of local service observations. A sample consisted of the calls observed at one central office in one month, and the series of samples used was selected from a complete record for all the central offices in a large city by the requirement that the number of calls per sample be not less than 450 nor more than 550. Distribution (b6) was obtained for the number of incorrect reports, (b7) for the number of cutoffs, (b8) for the number of double connections, and (b9) for the number of calls for the wrong number.

Group three is headed by Bortkewitsch's classical example of the Poisson exponential.[19] He found from the records of the Prussian army the number of men killed by the kick of a horse in each of 14 corps in each of 20 successive years, and, after discarding the records for 4 corps which were considerably larger than the others, treated the rest as one series of samples. This is distribution (c1). Series (c2) is similar to (b4), except that the samples of 100 two-place numbers were obtained from several different sources, logarithmic tables, trigonometric tables, and numbers listed in a telephone directory. Examples (c3), (c4), (c5), and (c6) show the variation in the number of telephone messages recorded per five-minute interval for certain groups of coin-box telephones in a large transportation terminal. The number of calls registered for each of 23 such telephones in each of about 20 five-minute intervals between noon and 2 p.m. was recorded on each of seven days (no Saturdays or Sundays included) but as the telephones are arranged in groups the distribution of the number of calls per interval was calculated for each group rather than for the individual telephones. These shown here are for a group of two telephones (c3), a group of four (c4), another group of two (c5), and a group of six (c6). The next four examples are

[18] "Logarithms, Their Nature, Computation, and Uses," by W. W. Duffield, Washington, 1897.

[19] Bortkewitsch, *op. cit.*

similar to examples (b6)–(b9), except that the limits of the number of calls per sample were 515 ± 25. Distribution (c7) was obtained for the number of cutoffs, (c8) for the number of double connections, (c9) for the number of calls for the wrong number, and (c10) for the number of connections to the wrong number. The next distribution (c11) was obtained from a count of the number of party-line subscribers listed per page of a large telephone directory and the last distribution of the group (c12) from a count of the number of advertisements in the "lost and found" column of *The New York Times* on each of the week-days from January 1, 1924 to August 31, 1924.

The fourth group contains only five examples, three of which are new. The first two of these present the same material used for example (b4) differently arranged. The 50,000 logarithms used are divided into 100 groups of 500 logarithms each for example (d1), and into 50 groups of 1,000 logarithms each for example (d2). The third (d3) is the distribution of the number of comets observed per year for the years 1789 to 1888 inclusive.[20] The other two distributions have been given by Perrin as typical of the data obtained when, in order to determine the density of the particles of an emulsion at a given depth, he restricted his field of vision to a tiny part of that layer, small enough so that the average number of particles visible was only one or two, and then made a large number of observations of the number of particles in that space at regular intervals.[21]

As was to be expected, these observed distributions have not only irregularities due to finite sampling but also in some cases what appear to be definite trends away from the corresponding Poisson distributions. In some cases there is an explanation ready at hand. For example, in series (b3), which gives the number of articles lost in the Telephone and Telegraph Building, the average number of articles lost per day might be expected to increase as the population of the building increased in this period following the completion of an addition, and the observed slant to the right is what would be expected. Also in series (d3), which gives the number of comets observed per year, the average would naturally increase steadily as a result of the continual improvement of telescopes and other instruments from 1789 to 1888. The curve toward the left in examples (c3) and (c5) might also be predicted because of the fact that the number of calls which could possibly be made in five minutes from a group of two telephones is certainly finite and probably rather small, and in examples (d4) and (d5) because it is difficult to judge by eye the number

[20] "Handbook of Astronomy," by G. F. Chambers, 4th ed., Oxford, 1889.

[21] "Brownian Movement and Molecular Reality," by Jean Perrin, London, 1910.

of particles visible simultaneously if that number is more than three or four.

In several cases special measures have been taken to reduce the variation of a and the resulting trend away from the corresponding Poisson distribution. In general, a is made as nearly constant as possible by making n and p constant throughout. In examples (b6)–(b9) and (c7)–(c10), for instance, each sample consists of approximately the same number of calls, and in example (c1) four corps were rejected because they were considerably larger than the others. In these examples it is assumed that p is practically constant and that by making n constant a constant average will be obtained. A somewhat different adjustment to keep a constant is illustrated by examples (a1) and (a2), where, as the decay of the radioactive substance decreases the average number of α particles emitted in a given solid angle per unit of time, the screen on which the particles strike is moved so that it intercepts a greater angle. In some cases n may be controlled much more easily than p, or vice versa, and a may be kept constant by letting one factor vary and adjusting the other to compensate, rather than by keeping both constant.

Summary

These examples of distributions which can be described by the Poisson exponential are of a dozen quite different kinds. They include eleven distributions found in published work on biometrics or statistics and twenty-one which are new. The agreement between the observed and the theoretical distribution is, in general, fairly good, and the applicability of the Poisson summation to a great variety of data is clearly indicated. The practical importance of some of these cases has been discussed above.

The use of the probability curves showing Poisson's exponential summation in place of double-entry tables as a source of data is shown to be simple, and their convenience as a background for plotting and comparing frequency-distributions is illustrated by Figs. 4 and 6–9. The new chart with a logarithmic scale for a (Fig. 5) is convenient in comparing distributions of different averages. It also shows the complete set of curves up to $a=30$ instead of only to $a=15$, and it makes it possible to read with considerable accuracy values of the variables in the range $0.1 \leqq a \leqq 2$, which is not clearly shown in Fig. 1 or Fig. 2.

TABLE II

c = number of occurrences of the event per sample.
m = number of samples with exactly c occurrences.
f = number of samples with at least c occurrences.
F = relative frequency of at least c occurrences per sample.

a1 *Alpha particles*
total = 10097
average = 3.87

c	m	f	F
0	57	2608	1.000
1	203	2551	.978
2	383	2348	.900
3	525	1955	.753
4	532	1440	.552
5	408	908	.348
6	273	500	.192
7	139	227	087
8	45	88	.034
9	27	43	.0165
10	10	16	.0061
11	4	6	.0023
12	0	2	.00077
13	1	2	.00077
14	1	1	.00038

a2 *Alpha particles*
total = 10094
average = 7.74

c	m	f	F
0	0	1304	1.0000
1	3	1304	1.0000
2	17	1301	.9977
3	46	1284	.9847
4	99	1238	.949
5	126	1139	.873
6	151	1013	.777
7	187	862	.661
8	180	675	.518
9	173	495	.380
10	131	322	.247
11	75	191	.146
12	44	116	.089
13	35	72	.055
14	16	37	.028
15	14	21	.0161
16	1	7	.0054
17	1	6	.0046
18	2	5	.0038
19	1	3	.0023
20	1	2	.00153
21	1	1	.00077

a3 *Deaths of aged*
total = 903
average = 0.82

c	m	f	F
0	484	1096	1.000
1	391	612	.558
2	164	221	.202
3	45	57	.052
4	11	12	.0109
5	1	1	.00091

a4 *Deaths of aged*
total = 2364
average = 2.16

c	m	f	F
0	162	1096	1.000
1	267	934	.852
2	271	667	.609
3	185	396	.361
4	111	211	.193
5	61	100	.091
6	27	39	.036
7	8	12	.0109
8	3	4	.0036
9	1	1	.00091

a5 *Telephone lines in use**
total = ?
average = 4.32

c	M	F
0	.013	1.000
1	.045	.987
2	.125	.942
3	.185	.817
4	.187	.632
5	.186	.445
6	.126	.259
7	.071	.133
8	.036	.062
9	.018	.026
10	.005	.008
11	.002	.003
12	.001	.001

a6 *Bacilli*
total = 1927
average = 1.93

c	m	f	F
0	219	1000	1.000
1	267	781	.781
2	219	514	.514
3	129	295	.295
4	70	166	.166
5	50	96	.096
6	26	46	.046
7	13	20	.020
8	5	7	.007
9	2	2	.002

b1 *Yeast cells*
total = 720
average = 1.80

c	m	f	F
0	75	400	1.000
1	103	325	.813
2	121	222	.555
3	54	101	.253
4	30	47	.118
5	13	17	.043
6	2	4	.0100
7	1	2	.0050
8	0	1	.0025
9	1	1	.0025

b2 *Yeast cells*
total = 1872
average = 4.68

c	m	f	F
0	0	400	1.000
1	20	400	1.000
2	43	380	.950
3	53	337	.843
4	86	284	.710
5	70	198	.495
6	54	128	.320
7	37	74	.185
8	18	37	.093
9	10	19	.048
10	5	9	.023
11	2	4	.010
12	2	2	.005

b3 *Lost articles*
total = 439
average = 1.04

c	m	f	F
0	169	423	1.000
1	134	254	.600
2	74	120	.284
3	32	46	.109
4	11	14	.033
5	2	3	.0071
6	0	1	.0024
7	1	1	.0024

b4 *Number 12*
total = 421
average = 0.84

c	m	f	F
0	231	500	1.000
1	150	269	.538
2	92	119	.238
3	24	27	.054
4	1	3	.006
5	1	2	.004
6	1	1	.002

b5 *Fires***
total = 9487
average = 26.1

c	m	f	F
0		364	1.0000
5		364	1.0000
10		363	.9973
15		346	.951
20		286	.786
25		185	.508
30		103	.283
35		53	.146
40		22	.060
45		18	.049
50		8	.022
55		4	.0110

b6 *Incorrect reports*
total = 138
average = 0.27

c	m	f	F
0	388	506	1.000
1	102	118	.233
2	12	16	.032
3	4	4	.0079

b7 *Cutoffs*
total = 1057
average = 2.09

c	m	f	F
0	75	506	1.000
1	126	431	.852
2	141	305	.603
3	73	164	.324
4	50	91	.180
5	29	41	.081
6	6	12	.024
7	2	6	.0119
8	3	4	.0079
9	0	1	.0020
10	0	1	.0020
11	1	1	.0020

b8 *Double connections*
total = 1760
average = 3.48

c	m	f	F
0	21	506	1.000
1	63	485	.958
2	98	422	.834
3	97	324	.640
4	85	227	.449
5	61	142	.281
6	42	81	.160
7	18	39	.077
8	11	21	.042
9	6	10	.0198
10	3	4	.0079
11	0	1	.0020
12	0	1	.0020
13	0	1	.0020
14	0	1	.0020
15	1	1	.0020

b9 *Calls for wrong number*
total = 2520
average = 4.98

c	m	f	F
0	10	506	1.000
1	20	496	.980
2	45	476	.941
3	60	431	.852
4	85	371	.733
5	92	286	.565
6	73	194	.383
7	55	121	.239
8	28	66	.130
9	18	38	.075
10	9	20	.040
11	5	11	.022
12	3	6	.0119
13	2	3	.0059
14	1	1	.0020

c1 *Deaths from kick of horse*
total = 122
average = 0.61

c	m	f	F
0	109	200	1.000
1	65	91	.455
2	22	26	.130
3	3	4	.020
4	1	1	.005

c2 *Number 12*
total = 251
average = 1.00

c	m	f	F
0	90	250	1.000
1	95	160	.640
2	46	65	.260
3	15	19	.076
4	3	4	.016
5	0	1	.004
6	0	1	.004
7	1	1	.004

c3 *Calls from group of two coin-box telephones*
total = 172
average = 1.19

c	m	f	F
0	44	145	1.000
1	48	101	.697
2	38	53	.366
3	13	15	.103
4	1	2	.0138
5	1	1	.0069

c4 *Calls from group of four coin-box telephones*
total = 384
average = 2.74

c	m	f	F
0	5	140	1.000
1	33	135	.964
2	24	102	.729
3	38	78	.557
4	23	40	.286
5	9	17	.121
6	4	8	.057
7	4	4	.029

c5 *Calls from group of two coin-box telephones*
total = 212
average = 1.50

c	m	f	F
0	27	141	1.000
1	49	114	.809
2	39	65	.461
3	19	26	.184
4	7	7	.050

c6 *Calls from group of six coin-box telephones*
total = 468
average = 3.39

c	m	f	F
0	8	138	1.000
1	13	130	.942
2	20	117	.848
3	37	97	.703
4	24	60	.435
5	20	36	.261
6	8	16	.116
7	5	8	.058
8	2	3	.022
9	1	1	.0072

c7 *Cutoffs*
total = 557
average = 2.09

c	m	f	F
0	44	267	1.000
1	62	223	.835
2	71	161	.603
3	43	90	.337
4	25	47	.176
5	14	22	.082
6	4	8	.030
7	2	4	.0150
8	2	2	.0075

c8 *Double connections*
total = 906
average = 3.39

c	m	f	F
0	14	267	1.000
1	33	253	.948
2	48	220	.824
3	56	172	.644
4	43	116	.434
5	34	73	.273
6	22	39	.146
7	8	17	.064
8	4	9	.034
9	3	5	.0187
10	2	2	.0075

c9 *Calls for wrong number*
total = 1351
average = 5.06

c	m	f	F
0	3	267	1.000
1	12	264	.989
2	23	252	.944
3	31	229	.858
4	45	198	.742
5	50	153	.573
6	37	103	.386
7	29	66	.247
8	13	37	.139
9	12	24	.090
10	4	12	.045
11	4	8	.030
12	3	4	.0150
13	1	1	.0037

c10 *Connections to wrong number*
total = 2334
average = 8.74

c	m	f	F
2	1	267	1.0000
3	5	266	.9963
4	11	261	.978
5	14	250	.936
6	22	236	.884
7	43	214	.801
8	31	171	.640
9	40	140	.524
10	35	100	.375
11	20	65	.243
12	18	45	.169
13	12	27	.101
14	7	15	.056
15	6	8	.030
16	2	2	.0075

c11 *Party lines*
total = 1981
average = 6.60

c	m	f	F
0	7	300	1.000
1	9	293	.977
2	14	284	.947
3	17	270	.900
4	21	253	.843
5	40	232	.773
6	46	192	.640
7	42	146	.487
8	32	104	.347
9	17	72	.240
10	22	55	.183
11	12	33	.110
12	6	21	.070
13	10	15	.050
14	1	5	.0167
15	3	4	.0133
16	0	1	.0033
17	1	1	.0033

c12 *"Lost and found" advertisements***
total = 7051
average = 33.7

c	m	f	F
0		209	1.0000
5		209	1.0000
10		208	.9952
15		207	.9904
20		199	.952
25		182	.871
30		144	.689
35		93	.445
40		51	.244
45		21	.100
50		7	.033
55		2	.0096

d1 *Number 12*
total = 421
average = 4.21

c	m	f	F
0	2	100	1.00
1	6	98	.98
2	18	92	.92
3	13	74	.74
4	16	61	.61
5	19	45	.45
6	13	26	.26
7	5	13	.13
8	5	8	.08
9	2	3	.03
10	1	1	.01

d2 *Number 12* total=421 average=8.42			
c	*m*	*f*	*F*
3	1	50	1.00
4	5	49	.98
5	2	44	.88
6	6	42	.84
7	6	36	.72
8	5	30	.60
9	7	25	.50
10	6	18	.36
11	4	12	.24
12	5	8	.16
13	1	3	.06
14	0	2	.04
15	2	2	·04

d3 *Comets* total=258 average=2.58			
c	*m*	*f*	*F*
0	19	100	1.00
1	19	81	.81
2	17	62	.62
3	14	45	.45
4	13	31	.31
5	8	18	.18
6	4	10	.10
7	2	6	.06
8	3	4	.04
9	1	1	.01

d4 *Particles in emulsion* total=46 average=0.92			
c	*m*	*f*	*F*
0	22	50	1.00
1	16	28	.56
2	7	12	.24
3	4	5	.10
4	1	1	.02

d5 *Particles in emulsion* total=106 average=2.12			
c	*m*	*f*	*F*
0	6	50	1.00
1	11	44	.88
2	12	33	.66
3	14	21	.42
4	6	7	.14
5	1	1	.02

* M is the relative frequency of exactly c occurrences per sample. Holm does not state the actual number of samples from which this was calculated, but it was evidently at least 1000.

** Since in the range $a>30$ the curves are drawn only for every fifth value of c, in these two distributions which extend beyond $a=30$ the values of f and F are tabulated only for every fifth value of c, and the values of m, which are meaningless unless the complete series is given, are omitted.

Line Current Regulation in Bridge Polar Duplex Telegraph Circuits

By S. D. WILBURN

SYNOPSIS: A mathematical analysis of the bridge polar duplex telegraph circuit, under the condition that the bridge arms are of equal resistance, shows that there is a particular bridge arm resistance which results in maximum received current. As the bridge arm resistances are increased beyond the value giving this maximum, the received current diminishes gradually. On the other hand, as the bridge arm resistances are decreased below the value giving the maximum, the received current drops off very rapidly. It follows that when necessary to limit line current, the maximum received current is obtained by placing the regulating resistance in the bridge arms. Also when the line resistance is large enough to limit the line current to less than the maximum allowable value, a gain may be obtained by increasing the bridge arm resistance to the value which corresponds to maximum received current. Experience has shown that in many situations where difficulty is encountered in operating a duplex telegraph circuit with the regulating resistances *in the line*, a very decided improvement is obtained by transferring these resistances to the bridge arms.

FOR the operation of polar duplex telegraph circuits, line batteries of uniform voltage are generally used and it is usually desirable to maintain the line current within fairly definite limits. The most suitable line battery voltage and the desired limits for the line current depend upon the type of line and apparatus used. In order to maintain the line current within the desired limits with uniform voltage it is necessary to add resistance to the circuit in greater or less amounts depending upon the length and gauge of the line circuit used. On account of line trouble and the necessity for rerouting telegraph circuits for other reasons, it is frequently desirable to switch a duplex set from one line to another of different resistance. To facilitate line current regulation without delaying service when such changes in line assignment are made, it is of considerable operating advantage to include in the wiring of each duplex circuit an adjustable resistance in the form of a rheostat mounted in an accessible location at the duplex set so that the attendant can readily regulate the line current at the time that necessary adjustments in the balancing artificial line are made to suit the changed line condition.

This paper outlines an investigation which was made with the object of finding an arrangement of line current regulating resistance which would result in the maximum steady-state received current with the bridge duplex telegraph circuit shown by Fig. 1, where it is desired to limit the line current to about .070 ampere. The condition for maximum steady-state received current was sought as the first step toward determining the most suitable arrangement of the resistances with the viewpoint that such an arrangement would probably be the most satisfactory from a transmission standpoint if it did not adversely effect the important factor of received current wave shape. An arrangement of the resistances was found which results in the maximum steady-state received current and from oscillographic tests which were subsequently made, this arrangement fortunately appears to improve the wave shape of the received current as compared with that resulting from other possible arrangements considered. It was also found from field trials on a number of practical circuits that this arrangement

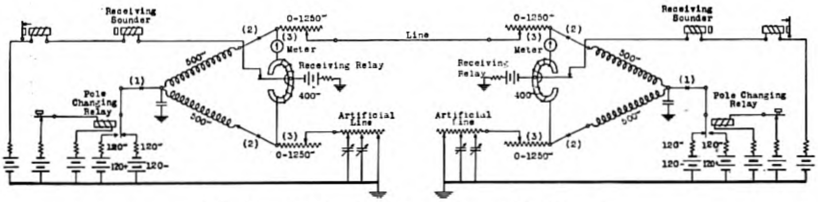


Fig. 1—Bridge Duplex Telegraph Circuit

results in improved transmission over other arrangements which have been considered for the regulating resistances.

Three different locations in the bridge duplex circuit are considered for the regulating resistances. These locations are designated (1), (2) and (3) in Fig. 1 and may be described respectively as follows:

(1) A single resistance in series with the battery branch of the circuit.

(2) Equal resistances in series with each of the bridge arms.

(3) Equal resistances in series with the line and the artificial line of the duplex set.

In considering locations (2) and (3), it is assumed that the resistances are in the form of a double rheostat with the movable arms mechanically connected to facilitate adding equal amounts of resistance simultaneously.

It will be seen from the circuit shown by Fig. 1 that of the three locations for the regulating resistance, (3) might be expected to reduce the received current most for a given line current, as that arrangement introduces resistance directly between the receiving relays. However, as that location for the resistances had been in general use, and since it was not at all obvious which of the other two arrangements would be the most favorable from the standpoint of received current, it seemed desirable to set up line current and received current equations to determine how the currents would be affected by the resistances in each location. Of the six current equations required, the one for the received current with the resistance in location (2) was found to possess a maximum within a resistance range which made that arrangement the most favorable from the standpoint of steady-state received current.

Curves i_1, i_2 and i_3 Fig. 2, show the steady-state value of received current which will be obtained on lines of 500 to 2260 ohms resistance

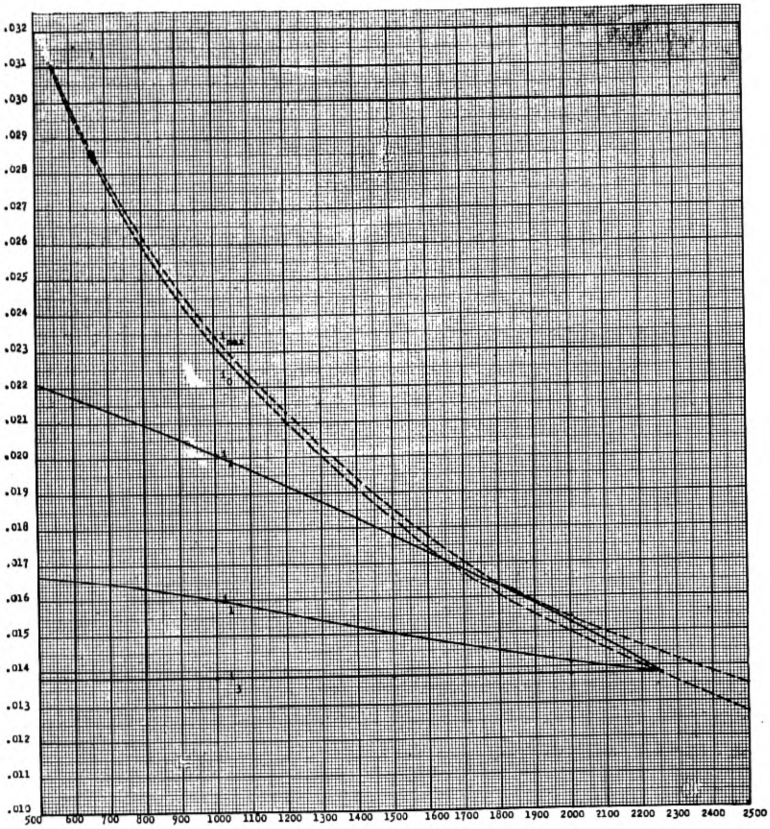


Fig. 2

with the regulating resistances located at points (1), (2) and (3), respectively. In each case just sufficient resistance is added to make the line current .070 ampere. If the resistance of the line is greater than 2260 ohms, the line current will fall below .070 ampere without the addition of resistance at either point. It will later be shown that, regardless of line current limitations, location (2) results in the maximum practicable steady-state received current in bridge duplex operation the bridge arms being of equal resistance.[1]

The method of calculating curves i_1, i_2, and i_3 Fig. 2, will be discussed presently along with certain other mathematical considerations.

In setting up the equations for the received current and line current, certain practical operating conditions of the circuit are assumed; first, that the circuit as a whole be kept symmetrical by using the same amount of regulating resistance at each station and second, that the duplex sets be maintained in a state of balance for direct currents. Line leakage will be neglected.

To express the line and received currents as direct, or explicit functions of the regulating resistances under the assumed condition of the circuit requires the use of unusually cumbersome equations which may to some extent be avoided in the early part of the solution without sacrificing accuracy. The complicated nature of the equations is due largely to the intricate relation between the regulating resistances and the overall network resistance of the duplex set from the terminal of the line to ground and, in turn, the relation between this network resistance and the two currents. With the exception of one step in the present investigation the work has been shortened by representing this network resistance by a parameter, or second independent variable, r which is itself a quadratic function of the regulating resistance, represented by R. The required values of r are then computed from the equation connecting it to R. In the expressions for the ratios of received current to line current all determinants which cannot be readily reduced to the second order cancel out so that considerable work is avoided by using these ratios rather than the explicit line current equations for calculating the line currents.

The equations expressing the relation between the received currents, i_1, i_2 and i_3 and the regulating resistance in the three locations (1), (2) and (3) respectively, are as follows:

$$i_1=\frac{aE}{(T+R_1+r_1)\,(2a+b)+ab+a^2} \tag{1}$$

$$i_2=\frac{R_2E}{(T+G+r_2)\,(2R_2+b)+bR_2+R_2^2} \tag{2}$$

$$i_3=\frac{aE}{(T+R_3+G+r_3)\,(2a+b)+ab+a^2} \tag{3}$$

[1] For the case of unequal bridge arms see Heaviside's "Electrical Papers," Vol. I, p. 24.

and the expressions for the ratios of the received currents to corresponding line currents, I_1, I_2 and I_3 are:

$$\frac{i_1}{I_1}=\frac{a(1/2T+r_1)}{(T+r_1)\,(2a+b)+ab} \tag{4}$$

$$\frac{i_2}{I_2}=\frac{R_2(1/2T+r_2)}{(T+r_2)\,(2R_2+b)+R_2b} \tag{5}$$

$$\frac{i_3}{I_3}=\frac{a\,(1/2T+r_3)}{(R_3+T+r_3)\,(2a+b)+ab} \tag{6}$$

where,

a, represents the constant resistance of each bridge arm in arrangements (1) and (3);

b, the resistance of the receiving relay;

E, the voltage of the line battery which is assumed to be equal at both stations and may be either negative or positive;

G, the constant resistance in series with the line battery taps in arrangements (2) and (3);

T, the resistance of the line between the duplex sets.

R_1, R_2 and R_3 are the regulating resistances in the different locations corresponding to the subscripts. In the equations for arrangement (1), G is assumed to be contained in R_1 and in the equations for arrangement (2), a is assumed to be contained in R_2. The equations for the parameters, r_1, r_2 and r_3 are as follows:

$$r_1=\sqrt{\tfrac{1}{4}T^2+R_1T+\frac{aT(a+b)+ab(a+2R_1)}{2a+b}}-\tfrac{1}{2}T \tag{7}$$

$$r_2=\sqrt{\tfrac{1}{4}T^2+GT+\frac{R_2T(R_2+b)+bR_2(R_2+2G)}{2R_2+b}}-\tfrac{1}{2}T \tag{8}$$

$$r_3=\sqrt{\tfrac{1}{4}T^2+GT+R_3(T+R_3+2G)+\frac{2aR_3(a+b)+aT(a+b)+ab(a+2G)}{2a+b}}-\tfrac{1}{2}T \tag{9}$$

While the line current and the received current can be calculated for any values of R and T from equations (1) to (9) inclusive, explicit line current equations are needed for calculating the received current for a definite value of line current, such as shown by curves i_1, i_2 and i_3, Fig. 2. It is clear that these curves cannot be calculated from equations (1) and (9) alone, as the first step necessary is to determine the value of R which, with a given value of T, will result in the specified value of I (.070 ampere). With line current equations R can, of course, be calculated by substituting .070 for I. While the line current equations can be set up fairly readily, they are of an extremely cumbersome character. For that reason curves i_1, i_2 and i_3, Fig. 2, were calculated by the following method:

From equations (1) to (9) inclusive, the line current was calculated for the various values of T from 500 to 3000 ohms with various values of R from 0 to 2000 ohms in steps of 250 ohms. For each value of T the line current was then plotted against R and the required value of the latter read from the intersection of the curve and the .070 ordinate. The values of R thus obtained were then substituted in equations (7) to (9) for calculating r. These values of R and r in turn were substituted in equations (1) to (3). By the above method the values of R within plus or minus two or three ohms can be determined. This possible error in R will not appreciably effect the points on the curves. The point of intersection of i_1, i_2, i_3, and i_0, Fig. 2 was calculated by equating the right hand side of equation (10) to .0138.

Referring to equations (1), (2) and (3) showing the relations between the regulating resistances and the received currents, it will be noted that in the right hand member of (1) and (3), R_1 and R_3, respectively, appear only as positive terms in the denominator. This shows that the received current will inevitably be reduced for every increase in the resistance, provided r_1 and r_3 are continuously increasing functions of R_1 and R_3 and from equations (7) and (9) it will be seen that both r_1 and r_3 increase continuously for every increase in R_1 and R_3, respectively. In equation (2), however, R_2 appears in both the numerator and the denominator and in the latter it appears in both the first and second powers. It is, therefore, not so easy to determine from an inspection of the equation just how the received current will be affected by increasing the resistance. It will be seen that this difference in the received current equations offers a guide in the selection of the location for the resistances which will result in the greatest received current.

From a closer inspection of equation (2), it is seen that when $R_2=0$ the received current will be 0 and, as the denominator of the right hand member contains the second power of R_2, the received current will approach 0 if R_2 be increased indefinitely. Also, it is clear that there will be current in the receiving relay for all finite values of R_2. Thus, if R_2 be indefinitely increased from 0, the received current will rise from 0 to a maximum value and then descend again toward 0.

This suggests solving for the value of R_2 corresponding to the point where i_2 is a maximum by differentiating equation (2) with respect to R_2 and equating to 0. The nature of the equation shows also that i_2 will have but one maximum. If the value of R_2 corresponding to maximum i_2 proves to be greater than 500 ohms, it will open up the possibility of increasing the received current by adding the regulating resistances at points (2), Fig. 1.

In calculating the line and received currents for different values of R_2 it is, of course, permissible to calculate separately corresponding values of r_2 and then substitute these values as constants in equation (2). Obviously this procedure cannot be followed in finding the derivative of i_2 with respect to R_2. The expression to be dealt with in this differentiation is that which results from the substitution of the right hand member of equation (8) for r_2 in equation (2). This substitution gives the following explicit and rigorous equation for the steady-state current in the receiving relays of a balanced symmetrical bridge duplex telegraph circuit:

$$i_2 = \frac{ER_2}{(\frac{1}{2}T+G)(2R_2+b)+R_2(R_2+b)+\sqrt{T(\frac{1}{4}T+G)(2R_2+b)^2+(2R_2+b)[R_2T(R_2+b)+bR_2(R_2+2G)]}} \qquad (10)$$

Equation (10) was found useful in calculating received currents as it combines (2) and (8) and may be used instead of equations (1) and (7) by changing G to R_1 and R_2 to a, but when it is differentiated and equated to 0 the resulting equation for R_2 corresponding to maximum received current is of an extremely impractical nature as it involves various powers of R_2 up to the sixth, together with an unusually large number of terms. In this investigation, it was not necessary to solve this equation for R_2 as it was found that for values of R_2 and T within the practical ranges of 500 to 1750 ohms for R_2 and 500 to 3000 ohms for T, r_2 is very nearly equal to 1/3 $R_2+2\sqrt{T}+200$. If this expression be substituted for r_2 in equation (2) and the result differentiated and equated to 0 it leads to the following equation which gives values of R_2 corresponding fairly close to the point of maximum received current:

$$R_2 = \sqrt{\tfrac{3}{4}b(T+2T^{\frac{1}{2}}+G+200)} \qquad (11)$$

With a receiving relay of 400 ohms resistance and a battery tap resistance of 120 ohms, as shown in Fig. 1, equation (11) becomes

$$R_2 = 10\sqrt{3(T+2T^{\frac{1}{2}}+320)}$$

From this equation, it is found that the bridge arms, each consisting of a 500 ohm bridge coil only, as shown by Fig. 1, are too small for maximum received current if the line resistance is greater than approximately 555 ohms. With line circuits ranging in resistance from 1000 to 2500 ohms, the respective values of R_2 necessary for maximum received current strength, range from approximately 624 to 935 ohms. If, then, resistance be added in the proper amounts at the points designated (2), Fig. 1, the received current will be increased thereby and at the same time, the line current will be reduced. If the line circuit resistance is approximately 1650 ohms or more the amount of resistance needed at points (2) to make the received current maximum, will be sufficient to reduce the line current to .070 ampere or less. This is illustrated by the three upper curves in Fig. 2. The lower broken curve, designated i_0, represents the received current which will be obtained with no regulating resistance in the circuit at either point. It will be seen that this curve passes below curve i_2 at a point corresponding to a line resistance of 1650 ohms. With approximately that value of line resistance and no regulating resistance, the line current is approximately .086 ampere and the received current is .0171 ampere. If approximately 410 ohms be added at points (2) the line current will be reduced to .070 ampere and the received current will remain at .0168 ampere. Curve $i_{\text{max.}}$ shows the maximum received current which can be realized by adding correct amounts of resistance at points (2). The upper curve touches curve i_2 at a point corresponding to a line resistance of 1850 ohms. That is, with a line resistance of this value, the regulating resistance required to reduce the line current to .070 ampere is just sufficient to bring the received current up to the maximum. For lines of this resistance or greater, the line current can be reduced to .070 ampere or less and at the same time the received current is increased. It will be seen from Fig. 2, that as compared to locations (1) and (3) for the regulating resistance, the advantage of location (2) from a steady-state received current standpoint, becomes greater with lines of low resistance and amounts to 32.3% and 60.1% respectively, with a line of 500 ohms resistance. On the other hand, the increase in received current due to arrangement (2), as compared to the condition of no regulating resistance, becomes greater with lines of higher resistance, as shown by the divergence of the i_2 and $i_{\text{max.}}$ curves, Fig. 2.

With line resistances in the lower range, the amount of regulating resistance needed to make the received current maximum will not be enough to bring the line current down to the desired value of .070 ampere. For example, with a line of 500 ohms resistance, the 500 ohm bridge arms are already too large by approximately 14 ohms and 1470 ohms will be required at points (2) to bring the line current down to .070. The bridge arms will then be 1484 ohms greater than needed for maximum received current. The question then arises as to why arrangement (2) results in the subtsantial received current gains with lines of low resistance, as shown by curves i_1, i_2 and i_3, Fig. 2. This part of the problem can best be solved by plotting equation (10). Fig. 3 shows this equation plotted for a 1200 ohm line. It will be seen that, from the

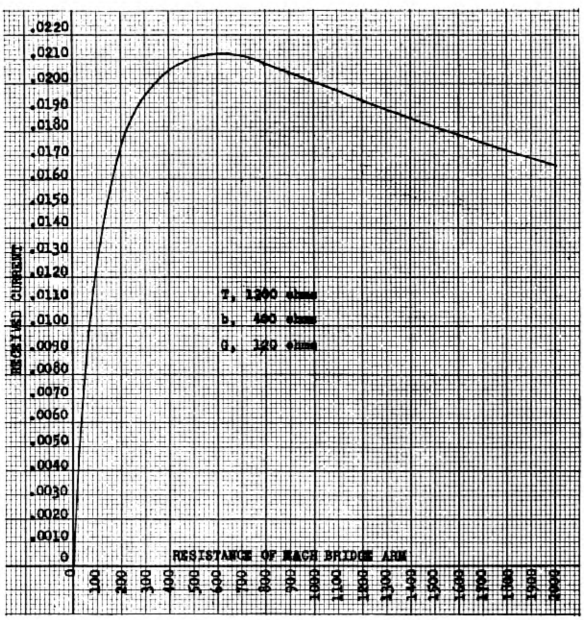


Fig. 3

standpoint of received current strength, it is better to have the bridge arm too great than too small, as the received current rises rapidly to a maximum and then descends slowly. On the other hand, if resistances be added at points (1) or (3), the operating point on the received current curve will in all cases be moved further away from the maximum, and this movement away from the maximum will take place on the side of the maximum which has the greatest effect in reducing the received current, as will be shown.

The resistance at points (1) or (3) moves the operating point on the received current curve away from the maximum due to the fact that the value of the bridge arm resistance corresponding to maximum received current is a function of both the resistance between the duplex sets and the resistance in the battery branch of the circuit, as shown by equation (11); G in this equation corresponds to R_1 in equation (1) and T represents all the resistance between the duplex sets, this being augmented by the addition of resistance at points (3). Therefore, any increase in the resistance in the battery branch of the circuit or in the resistance between the sets, such as will result

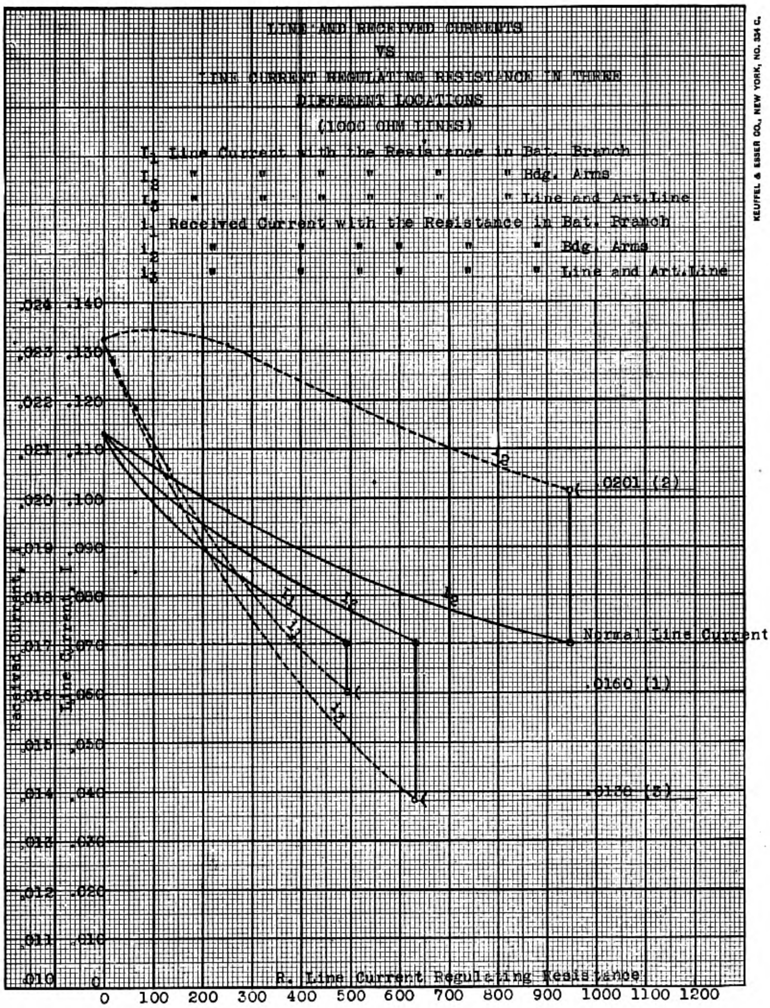


Fig. 4—Line and Received Currents vs Line Current Regulating Resistance in Three Different Locations—(1000 Ohm Line)

from adding resistance at points (1) or (3) respectively, will move the maximum point on the curve further to the right. This can best be illustrated by the following example:

With a line of 1500 ohms resistance, the resistance required in each bridge arm for maximum received current is about 750 ohms, so that the normal 500 ohm bridge arms as shown by Fig. 1 are short of the maximum by 250 ohms. If the line current be reduced to the desired value of .070 ampere by adding resistance at points (3) about 250 ohms will be required at each station. This will make the resistance between the duplex sets, corresponding to T in equation (11), 1500+500=2000, for which the value of the bridge arms for maximum received current is about 855 ohms. The operating point on the curve is, therefore, 355 ohms on the left hand side of the maximum, as compared with 250 ohms before the resistances were added. The change in the maximum due to adding resistance at points (1) takes place in the same general way though not in exactly the same degree.

Fig. 4 shows how the line and received current are affected by the resistances in each location with a 1,000 ohm line. From these curves it will be noted that location (2) for the resistances results in a gain of about 25.6 per cent. in received current strength as against location (1) and as compared to location (3) the gain in received current amounts to about 45.6 per cent.

As the ratio of the bridge arms is not changed by adding the line current regulating resistance in equal amounts at points (2) that arrangement should introduce no difficulties in maintaining a balance between the line and artificial line. Furthermore, arrangement (2) should not increase disturbances due to small extraneous currents in the line.

Carrier-Current Communication on Submarine Cables

Los Angeles-Catalina Island Telephone Circuits [1]

By H. W. HITCHCOCK

SYNOPSIS: Seven telephone channels and one telegraph channel on one single-conductor deep-sea cable have been made possible by the employment of carrier current on one of the two submarine cables across Catalina channel. This is the only application of carrier telephony to deep-sea cables and the system is one of the shortest carrier systems (26 mi.) in commercial operation; it provides more separate carrier channels (six) than has been previously attempted; and it differs in other important respects from other systems. This paper describes this carrier-current system.

IN the commercial application of new developments in the electrical communication art, there are a few places which repeatedly call attention to themselves. Notable among these is Catalina Island, for it is probable that in providing telephone service across the short expanse of water which separates Catalina from the mainland, more novel improvements have been employed than at almost any other point.

The first commercial telephone communication with Catalina Island was established in 1920 when a radio system was placed in operation between Avalon and the mainland, the circuit being extended by wire to Los Angeles. This circuit was in use for several years and featured in a number of transcontinental demonstrations, including the one which was held at the opening of the service to Havana over the Key West-Havana cables.

The system is of considerable interest as it represents the only instance in which radio has been used, in this country at least, to form a portion of a toll telephone system for the general use of the public. That it was reasonably successful is demonstrated by the fact that on some days as many as 183 commercial telephone messages and a large number of telegrams were handled over it. The system also proved to be one of the first popular broadcasting stations and many letters were received from radio fans, often several hundred miles away, telling of some of the amusing conversations which were overheard.

In 1923 the radio was replaced by two single-conductor submarine cables. By that time the demands for service were too great to be met by a single circuit, while the growing interest in radio broadcasting, as well as the increasing interference from ship transmitters,

[1] Presented at the Pacific Coast Convention of the A. I. E. E., Salt Lake City Utah, Sept. 6–9, 1926.

rendered its continued operation very difficult and unsatisfactory. The submarine cables were of the single-conductor, deep-sea type, each providing a single-wire circuit. They are of interest for a number of reasons, chiefly, perhaps, because they represent one of the few instances of deep-sea cable manufacture in this country. From the cable hut at San Pedro, the circuit is extended to the office by means of a special lead-covered cable containing four individually shielded No. 13 B & S gauge pairs for the telephone circuits and four 19-gauge pairs for the telegraph circuits and other miscellaneous uses. Between the San Pedro office and Los Angeles, the circuit was composed of a No. 19 B & S gauge cable phantom. At San Pedro a through-line repeater was inserted in order to secure the desired over-all equivalent between Avalon and Los Angeles.[2]

Although the two circuits provided by the cables represented a great improvement over the previous condition as regards the quality of the service rendered and the number of messages which could be handled, it was realized that they would soon prove inadequate to handle the heavy summer business, for which eight or ten circuits would be required in a relatively short time. To provide for such a large increase by the laying of additional cables was deemed impracticable, as the cost would be excessive. Furthermore, in water of this depth—3,000 feet—it is important that cables be laid at least a mile or two apart, so that in the event that trouble develops on one, it can be repaired without disturbing any of the others. For a total distance as short as the width of the Catalina channel—23 nautical miles—such a separation between adjacent cables could not be maintained without materially increasing the length of the outer ones with a corresponding increase in their cost and in their transmission equivalents. In view of these facts, it was decided to secure as many more circuits as possible by operating carrier systems over the two cables already in use. This project was actively promoted with the result that on May 15, 1926, six carrier telephone circuits were placed in operation.

The use of carrier in the past few years has increased so rapidly that the mere addition of a new system is, in itself, of hardly more than passing interest. In this instance, however, there are a number of factors which render the project of particular interest. It is one of the shortest carrier systems—26 miles—in commercial operation. It is the only application of carrier telephone to deep-sea cables; the system pro-

[2] A description of these cables and their laying was given in a paper presented by the writer at the Pacific Coast Convention in 1923 and published in Volume XLII of the *Transactions.*

vides more separate channels (six) than has ever before been attempted, while the particular arrangement employed is different in many other important respects from anything which has been used in the past.

In order to better appreciate the reasons for adopting the system finally agreed upon, it may be of interest to review briefly the essential characteristics of carrier systems and the different types which are available.[3]

Carrier systems may be divided into two general classes, namely, balanced or grouped, depending upon the manner in which the currents in the two directions are prevented from interfering with each other

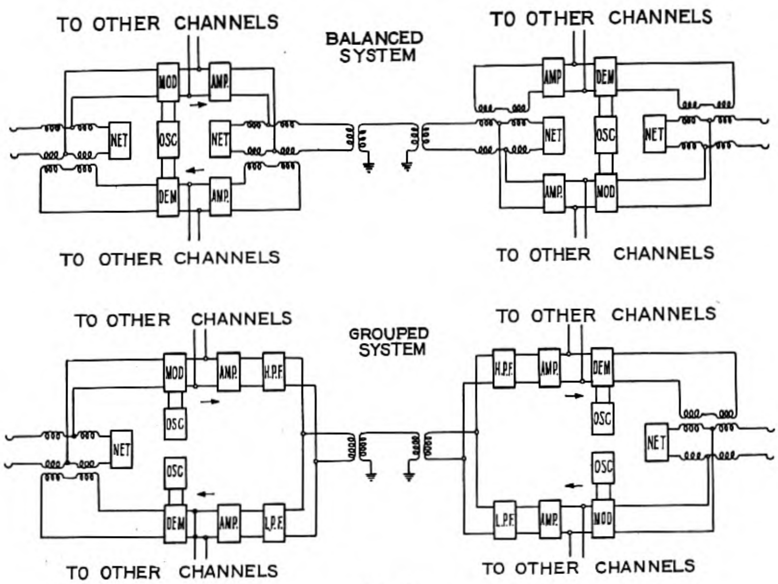


Fig. 1

at the terminals. In the balanced system this separation is accomplished by means of a three-winding transformer or hybrid coil together with a balancing artificial line such as is used with a voice-frequency repeater. In the grouped system, different carrier frequencies are used for transmission in the two directions and their separation at the terminals is effected by means of suitable band-pass filters. These two systems are shown diagrammatically in Fig. 1. The balanced system has the advantage that for each channel

[3] The general principles of carrier-current telephony are described at considerable length in a paper by Messrs. Colpitts and Blackwell which was published in Volume XL of the Journal of the Institute.

the same carrier frequency may be used for transmission in both directions so that there may be as many channels as there are separate carrier frequencies. On the other hand, the wire circuit must be very uniform throughout so that the impedance will be very regular over the entire carrier-frequency range, and may be simulated by an artificial line. The line must also be very stable so that the impedance balance, once having been secured, will not be disturbed. Furthermore, as transmission with the same carrier takes place in the two directions, the effect of the cross-talk between systems of the same type is very severe, so that it is usually impracticable to operate two of these over wires which are in close proximity for any considerable distance. The grouped system has the advantage that a balancing line is not required and hence small circuit irregularities are relatively unimportant. Furthermore, the effect of cross-talk is much less severe, so that a number of systems may often be operated over adjacent circuits. One disadvantage is that two carrier frequencies are required for each channel so that fewer circuits can be secured with one system.

Carrier systems may also be divided into two classes depending upon the manner in which the carrier current is provided at the receiving end. In the carrier transmission system, the carrier current is supplied by the oscillator at the sending end and is transmitted over the circuit along with one or both of the side bands. In the carrier suppression system, the carrier current itself is not transmitted but is introduced into the receiving equipment from a local source. This latter system is proving to be superior for general carrier purposes because of the advantages which accrue from relieving the line and apparatus from the load of the carrier current.

Turning now to the electrical characteristics of the cables, we find that each one provides a circuit having a transmission equivalent which increases throughout the carrier range but is moderate in magnitude. The impedance, as is to be expected with a uniform, non-loaded cable, is very smooth, and since there is no opportunity for any change in the cable constants, the impedance has practically no variation. The transmission equivalent and the impedance of one of the cables are shown in Figs. 2 and 3, respectively. The cross-talk between the cables is small enough to be entirely negligible, regardless of the type of carrier systems employed.

In view of all the conditions outlined, a balanced system of the carrier suppression type was decided upon. Such a system provides the maximum number of channels per cable, while the usual difficulties of impedance balance and inter-system cross-talk are largely absent due to the unusual characteristics of the cables. The adoption of such a system also made possible the employment of standard units of equipment of the most recent design. The general nature of the system and the arrangement of the component parts is shown diagrammatically in Fig. 4. Fig. 5 is a simplified circuit diagram

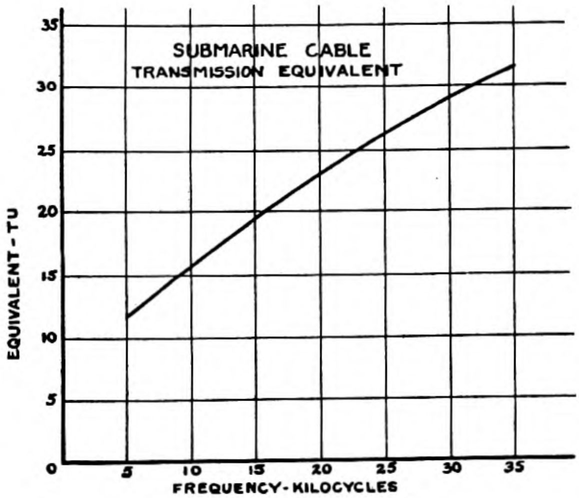


Fig. 2

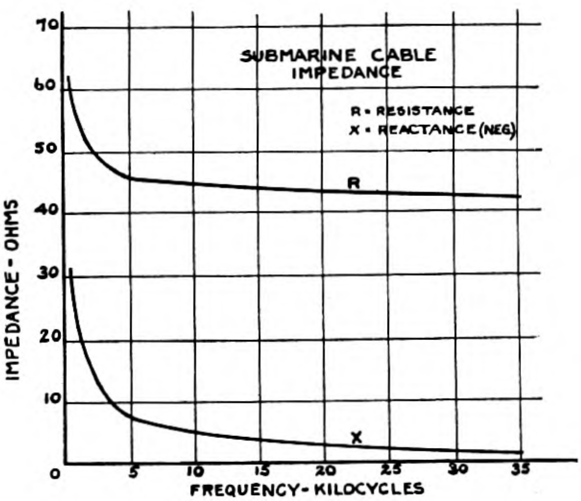


Fig. 3

showing the filters for separating the various circuits at the terminals, together with the balancing arrangement. In Fig. 6 are shown the essential parts of one channel together with the amplifiers and the hybrid coil which are common to all the channels. For convenience, some of the battery and auxiliary circuits have been omitted in the figure.

At the time the system was under development, it was uncertain that balanced operation of all channels over a single cable would be practicable, so that an alternative arrangement involving substan-

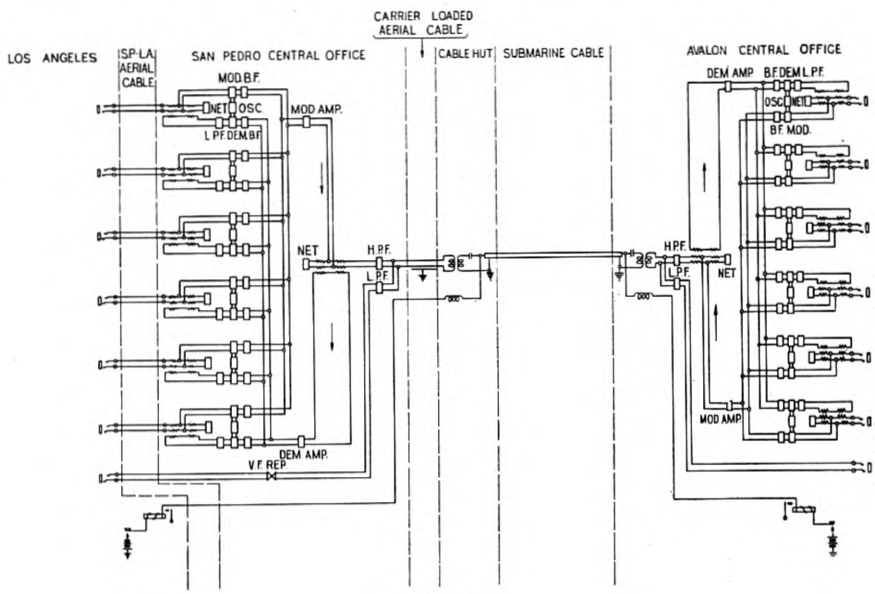
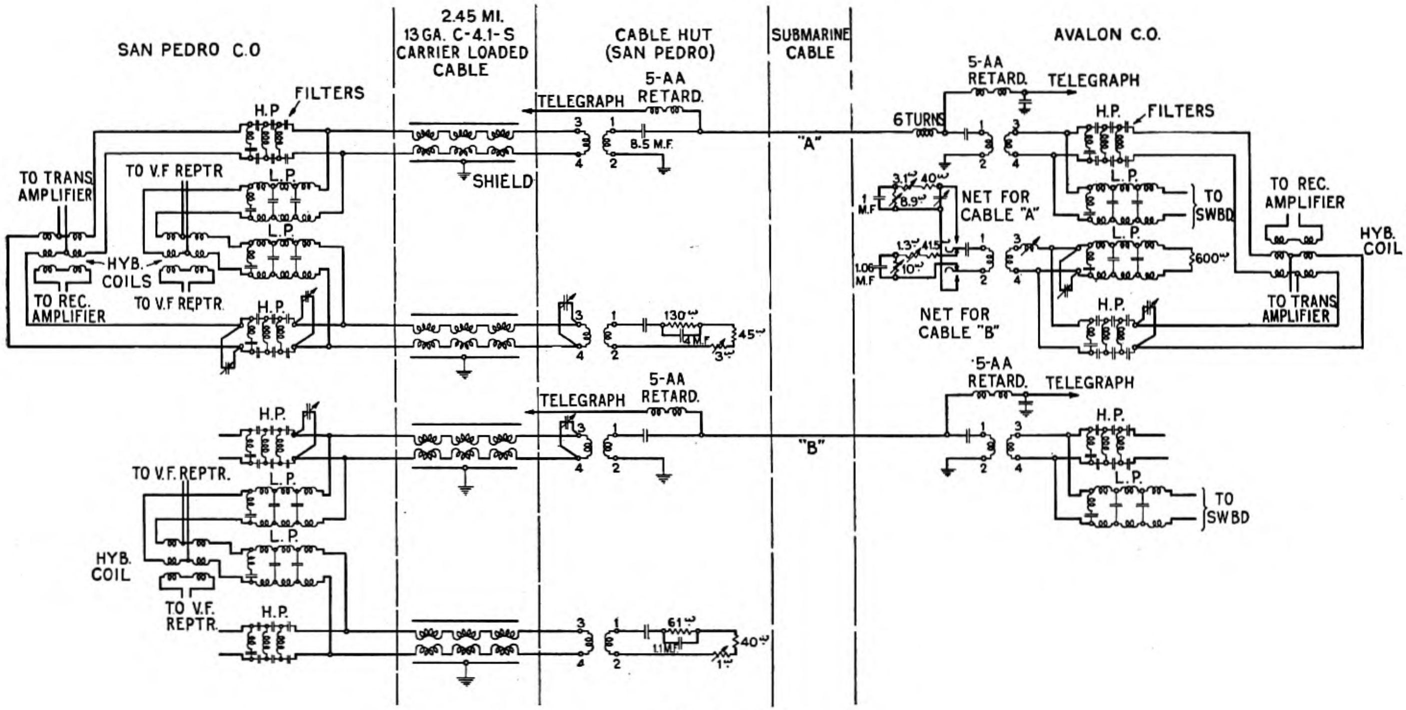


Fig. 4

tially four-wire operation over the two cables was provided for. With this arrangement, which is shown diagrammatically in Fig. 7, all transmission in one direction takes place over one cable, while transmission in the opposite direction is effected over the second cable. No balancing equipment or hybrid coils are employed. Such an arrangement would increase the system stability, if such were required, but would limit the total carrier capacity of the two cables to six channels. In the event of the failure of one cable, operation with such a system would be impossible, and it would be necessary, at that time, to revert to the two-wire arrangement as described above, with a possible reduction in the over-all gain or a reduction in the number of operating channels.



SAN PEDRO C.O
2.45 MI.
13 GA. C-4.1-S
CARRIER LOADED
CABLE
CABLE HUT
(SAN PEDRO)
SUBMARINE
CABLE
AVALON C.O.
FILTERS
H.P
L.P.
TO TRANS
AMPLIFIER
TO V.F REPTR
HYB.
COILS
TO REC.
AMPLIFIER
TO V.F REPTR.
SHIELD
TELEGRAPH
5-AA
RETARD.
8.5 M.F.
"A"
"B"
6 TURNS
NET FOR
CABLE "A"
NET FOR
CABLE "B"
TO
SWBD
600
HYB.
COIL
TO V.F. REPTR.
TO V.F.
REPTR.

Fig. 5

As may be seen from Fig. 4, a carrier-equipped cable provides a d-c. telegraph circuit, and one voice-frequency and six carrier-frequency telephone channels. The separation of the various channels is effected by means of electrical filters. Fig. 8 shows the band of frequencies employed for each channel. For the d-c. telegraph this

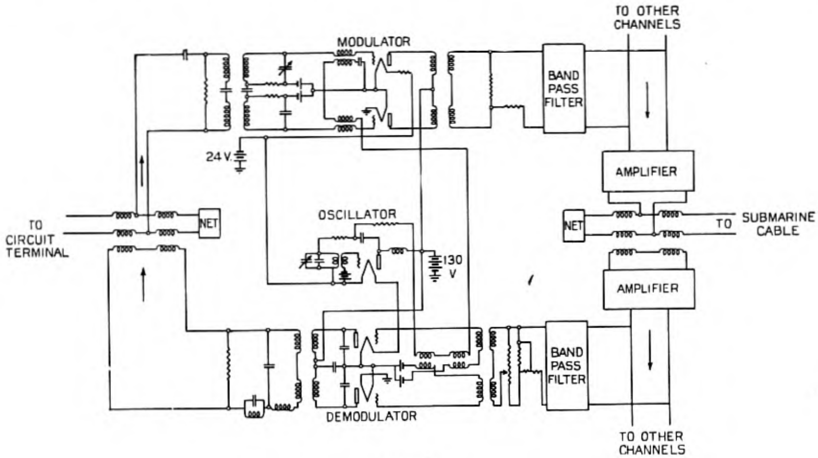


Fig. 6

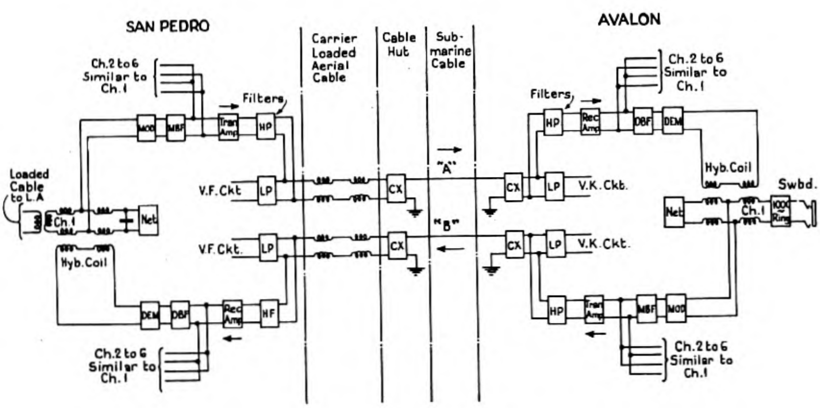


Fig. 7

separation is effected at the terminals of the cable as is shown in Fig. 5. The telegraph circuit requires a continuous d-c. path, whereas the telephone channels require the insertion of an inequality ratio insulating transformer at the ends of the cable in order to properly join the 43-ohm grounded cable circuit with the 600-ohm metallic circuit formed by the office equipment and intermediate cable. As this transformer must pass both the voice and carrier channels, it has been designed so as to have a high efficiency for all frequencies between 250 and 30,000 cycles. Separation of the voice-frequency circuit from the carrier system is performed by means of the usual high and low pass filters which are located at the central offices. These filters both have a cut-off frequency of 3,000 cycles, the low

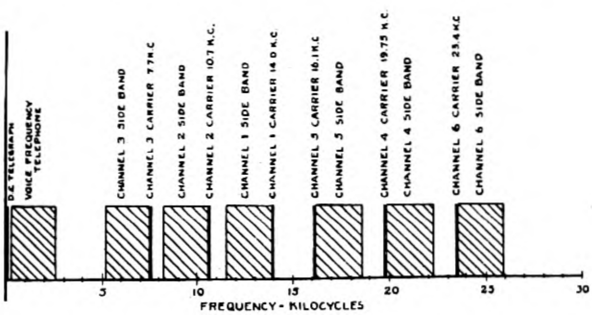


Fig. 8

pass transmitting all frequencies below this value and the high pass transmitting all above it. In the carrier system the transmitting and receiving currents are separated from each other by a hybrid coil and balancing network. Between the output of the six modulators and the common transmitting amplifier, individual band-pass filters are located. Each one of these filters is designed to transmit one of the side bands produced by the modulator associated with it and to suppress all other frequencies. The six receiving currents are separated in a similar manner. Each filter allows current of the proper frequency to pass to the corresponding demodulator and excludes all others. Each demodulator is also provided with a low pass filter which allows the passage of the resulting voice-frequency current but excludes all incidental higher frequencies which might be present and render the circuit noisy. The input of each modulator and the output of the corresponding demodulator are finally joined by means of a voice-frequency hybrid coil and extended to the circuit terminal as a two-wire circuit. At the San Pedro end, each two-wire circuit is extended to Los Angeles over a loaded cable circuit. Phantoms are employed for this purpose as they have a higher cut-off frequency than have the side circuits, with a correspondingly better quality.

Concerning the carrier system itself, the two ends are practically identical while the general equipment arrangement for an individual channel is the same in all cases except for the frequency of the band-pass filter. For this reason, a consideration of one channel is sufficient. Each channel is composed of a voice-frequency hybrid coil, a modulator with its band-pass filter, an oscillator, and a demodulator, together with its associated filters. In addition, there is, at each end, a carrier hybrid coil together with transmitting and receiving amplifiers which are common to all channels. The arrangement of this equipment is shown schematically in Fig. 6, as previously indicated.

The modulator, the input of which is connected to the center taps of the hybrid coil line windings, utilizes of two vacuum tubes arranged for push-pull operation. The carrier current which is supplied by the oscillator is applied to the two grids by means of a transformer. Such a circuit generates the two side bands but suppresses the carrier. In order that this suppression may be as complete as possible, the small condenser associated with the grid of one of the tubes is made variable and is adjusted until the carrier current in the modulator output is reduced to a minimum. The band-pass filter transmits one of the side bands and suppresses the other, as well as all miscellaneous resultant currents of a higher order which are produced by the modulator. It also prevents the output currents of the other channels from entering the modulator circuit as this would cause a reduction in their efficiency and give rise to undesirable frequencies.

The demodulator is very similar to the modulator. The tube arrangement is substantially the same and carrier current is supplied from the one oscillator. In the demodulator a complete suppression of the carrier is unnecessary as this is accomplished by the low pass output filter. For this reason, the small balancing grid condensers are omitted. In order to adjust the over-all gain of the channel, the demodulator is provided with an adjustable potentiometer graduated in two transmission unit steps, and in addition, fixed pads are provided for making further gain adjustments. The output of the demodulator is connected to the series winding of the voice-frequency hybrid coil.

The oscillator which supplies the carrier current to the modulator and demodulator is of the usual type. The tuning condenser includes a small variable unit for making small adjustments in frequency. Separate oscillators are used at the two ends for each channel, and as these are in no way connected together, it is occasionally necessary to make slight adjustments in order to keep the frequencies at the two ends substantially equal. The oscillators are very stable, however, and such adjustments are seldom required.

The individual channel filters are all of the band-pass type as previously indicated and have a free transmission range of approximately 2,500 cycles. Outside this free range they have a high impedance so as not to act as a shunt for the other channels. They are all of substantially the same construction, although the constants of the component parts necessarily vary as the filters for the different channels transmit different frequencies.

The transmitting and receiving amplifiers, which are practically identical, are shown schematically in Fig. 9. They consist of two push-pull stages connected in tandem. Each half of the second or output stage consists of two parallel tubes of high output capacity. In this way a comparatively high gain and a large energy output

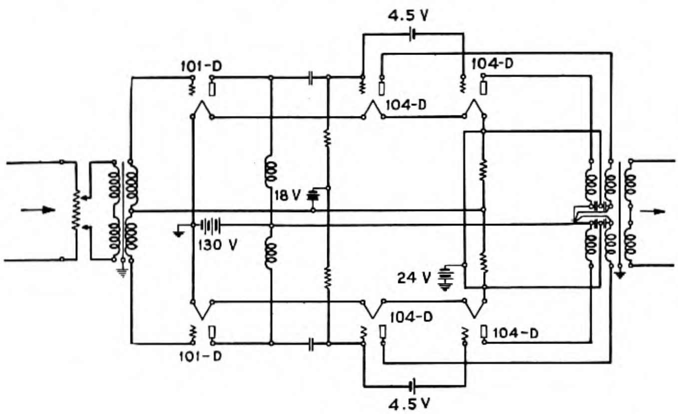


Fig. 9

may be secured without overloading. This is very important as these amplifiers are common to all six channels and any tendency to overload would produce objectionable distortion and inter-channel modulation. In order to adjust the over-all gain for the entire system, each amplifier is provided with an input potentiometer.

As has been previously indicated, the transmitting and receiving circuits are joined to the cable by means of hybrid coils. Probably the most difficult problem encountered in the installation of this system was the securing of an adequate balance. The difficulty of doing this may be better appreciated when it is realized that this balance must cover all frequencies from 3,000 to 30,000 cycles, and must have a value of from 30 to 45 T. U., the higher value which represents an impedance unbalance of approximately one per cent. being required at the upper frequency. In order to secure such a balance, every part of the line circuit was matched by a similar part in the network circuit. All filters and transformers on the line side of the hybrid coil were duplicated in the network, and on the San Pedro side a 13-gauge carrier-loaded cable pair was included in the network circuit between the office and the cable hut, and the inequality ratio transformer and basic network simulating the cable were located at the latter point. In addition to providing a balance within the carrier range, it was necessary at the San Pedro end for the network circuit to balance the cable within the voice-frequency range as a through-line repeater is employed on the voice-frequency circuit. Not only was it necessary to duplicate all parts in the line

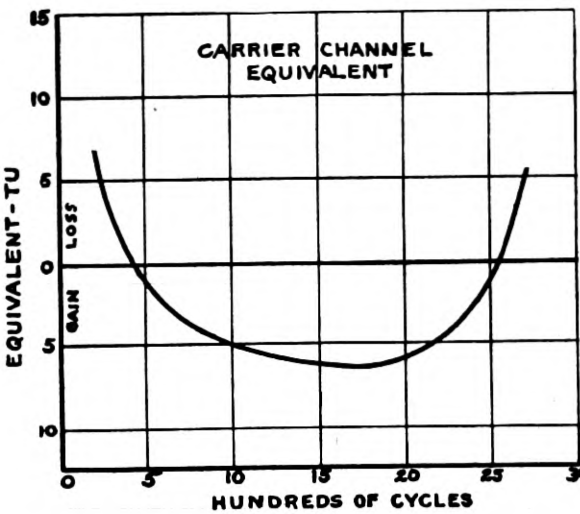


Fig. 10

and network circuits but in addition they were carefully selected and paired so that the two parts associated would have, as nearly as pcssible, the same electrical characteristics. All wire pairs within the office which appeared in the carrier frequency circuits were individually shielded by means of a grounded metallic covering. The 13-gauge carrier-loaded pairs in the cable joining the hut and the office were also individually shielded by means of a lead foil wrapping. This was done in order to preserve the balance and prevent cross-talk with another system which may be placed on the second cable at some future time.

Although extreme care was exercised in making the refinements described, the balance was still lower than was desired so that small variable auxiliary impedances were inserted at suitably chosen points in the line and network circuits. By the adjustment of these elements, it was found that the balance could be raised to any desired value for any particular channel, but that in so doing, the balance on some of the others would be impaired. By careful adjustment, however, it was possible to secure a balance for all channels within the range previously mentioned. As the transmission equivalent of the cable

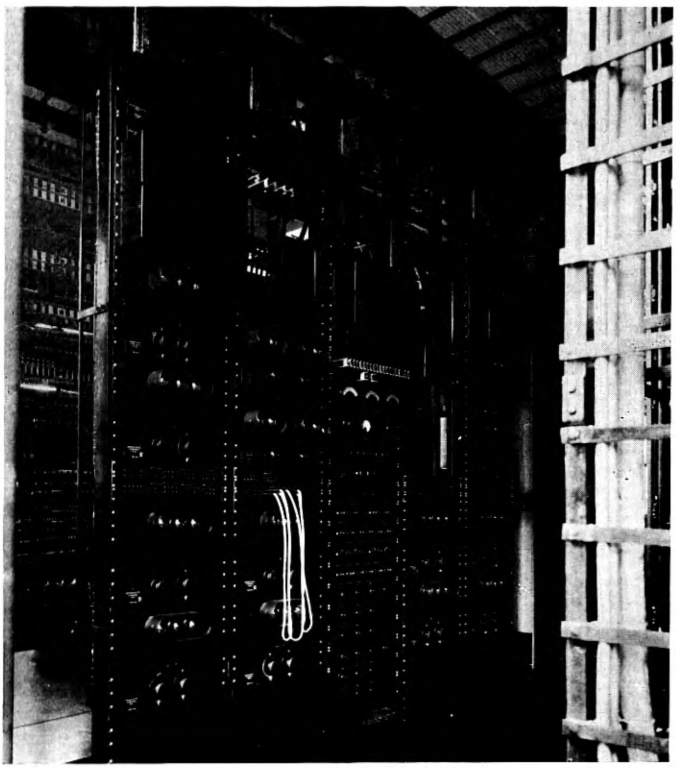


Fig. 11

increases with the frequency, the over-all channel gains must be increased in the same manner in order that all circuits may have the same over-all equivalent. The networks were therefore arranged so that the higher frequencies would have the better balance, as in that way the margin of balance over gain could be made substantially the same for all channels. Since this margin should not be allowed to fall below a fairly definite minimum if the circuit is to have the desired stability, it is evident that the balance which may be secured determines the over-all gain which is possible. In this case the circuit equivalent for all channels between Los Angeles and Avalon was set at five T. U. As the loaded cable between Los Angeles and San Pedro is approximately nine T. U., it may be seen that the carrier system actually introduces a gain and performs the function of a

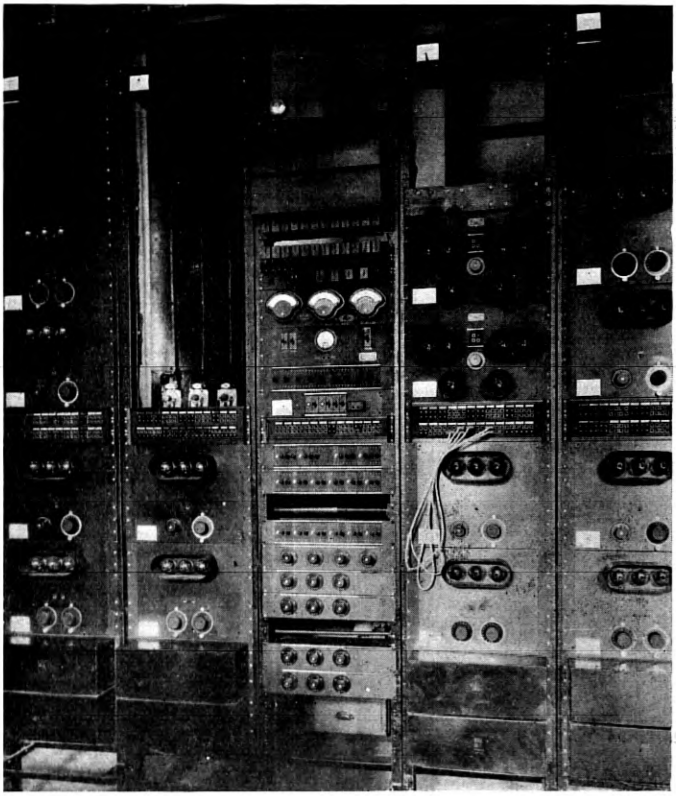


Fig. 12

repeater besides increasing the number of circuits. Fig. 10 gives a frequency characteristic of one of the channels which is typical of all of them. Balancing equipment has been provided for both cables as is shown in Fig. 5. With this arrangement, the carrier system may be operated over either cable. The transfer from one cable to the other is so simple that it can be made with practically no traffic interruption.

Signaling over the carrier channels is effected by means of 1,000-cycle ringers which are connected to the circuits at the two terminals. As the ringing current is within the voice range, it is transmitted over the regular carrier channel so that no additional signaling equipment is necessary.

In order to insure satisfactory operation, all necessary testing facilities are included. Meters and keys are provided for measuring the voltages of the plate, grid and filament batteries as well as the

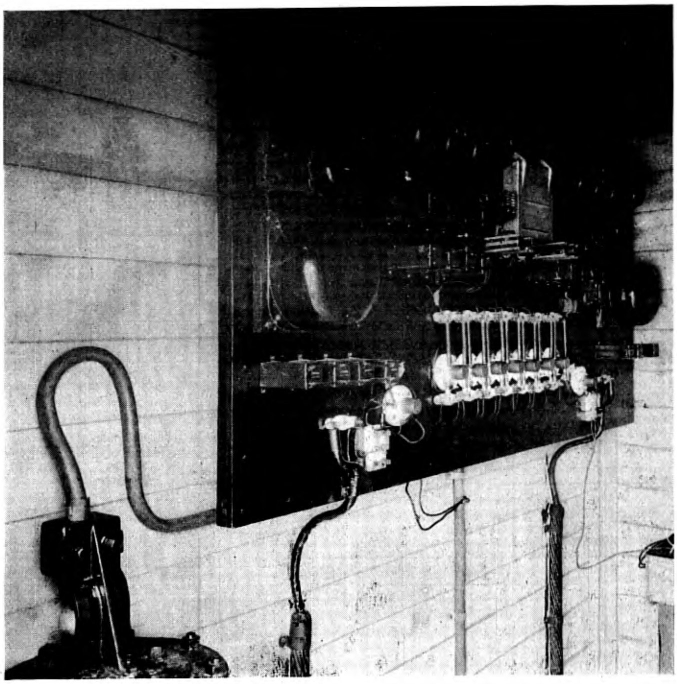


Fig. 13

plate and filament currents of all tubes. Individual rheostats are inserted in all filament circuits for making any adjustments that may be necessary. Alarms are provided to indicate any abnormal condition which might develop on any tube. Thermocouples and artificial lines have been conveniently arranged for checking the efficiency of all units such as the modulators and demodulators. Jacks are located at suitable points so that any changes which may be necessary can be quickly made.

The general appearance of the carrier system may be seen from Figs. 11 and 12 which show the equipment at San Pedro and Avalon respectively. Fig. 13 is an interior view of the San Pedro cable hut showing the cable terminals, together with the insulating transformers, telegraph composite sets, and basis networks. Referring to the central office equipment, the first bay contains the equipment for two complete channels. At the top are the terminal strips for making all connections with the equipment below. On the next two small panels are mounted the hybrid coils and the other miscellaneous apparatus associated with the voice-frequency ends of the two channels. Below these are the modulator and demodulator band filters which are covered with dust proof cases. Next comes the modulator and demodulator panels for one channel. Below the two jack strips is mounted similar equipment for a second channel but arranged in reverse order. In the upper half of the second bay is located a small panel mounting the carrier hybrid coil and associated equipment. Below this appear the transmitting and receiving amplifiers. The lower half of the bay is similar to the lower half of the first one. In the third bay is mounted all the battery supply and testing apparatus. The first two units contain the battery retard coils. Below these are the alarm relays and auxiliary resistances. Next come the meters for measuring the tube currents and voltages, and below these are the thermocouples and artificial lines for making high frequency measurements. Below the jack strip are the keys for opening and closing the individual filament circuits used for measuring the plate and filament currents. Alarm lamps are also associated with each of the filament circuits. The fourth bay is similar to the second except that the upper half is vacant. As may be seen from the photographs, the amplifiers appear on the second bay at San Pedro and on the fourth at Avalon. The fifth bay is an exact duplicate of the first.

The new system has now been in successful operation for the past five months. In the light of its performance thus far, we feel assured, that when more circuits are required a second system of six channels can be added to the second cable, thus providing a total of fourteen telephones and two telegraph circuits over the two single-conductor cables. Such a circuit group, we believe, will meet the traffic requirements for quite a number of years.

Abstracts of Recent Technical Books and Papers from Bell System Sources

Theory of Vibrating Systems and Sound. I. B. CRANDALL. Recent years have witnessed a great revival in Acoustics, both as a subject of industrial research and as a field for academic cultivation. This development has been carried on most actively in America and in Germany, and has been measurably due to the demands of the public for commercial acoustic devices in the public address and radio fields. In addition, there has evolved the new subject of Architectural Acoustics, largely through the researches of the late W. C. Sabine and his followers in this country.

In this situation, with an ever increasing body of technical literature, it may be noted that the standard textbooks on Sound have maintained their classical character, although new impedance methods and terminology have come into use (closely parallel to those of electrical theory) and many new fields of study have been opened up. The need for a connected treatment involving the new subject matter and methods the present author has attempted to supply, and while necessarily building on the classical treatises as a foundation, he has included a chapter on complex vibrating systems with a treatment of acoustic filters; two chapters on the theories of sound generation and radiation with applications to tubes and horns and a chapter on the essentials of architectural acoustics and absorbing materials. An extended bibliographic appendix serves as an entry to many branches of Applied Acoustics—for example, loud speaking telephones, piezo-electric resonators, recent work on speech and hearing, and submarine signalling, to cite only a few examples.

The author has purposely avoided duplicating classical material such as the theories of bars or of Fourier's series, feeling that the newer ideas and developments deserve the emphasis gained thereby.

The book aims to organize Acoustics as now practiced, and research workers will probably find it useful, not only in their classes, but as a starting point for acoustic research.

The book is produced by D. Van Nostrand & Co. ($5.00).

Contemporary Physics. KARL K. DARROW. During the twenty-five years since this century began, the science of physics has undergone amazing enlargements and transformations which may well be ranked among the most significant attainments of our times. Through discoveries and improvements in the arts of experimentation, it became possible to measure the charge and the mass of the atom of negative electricity; to measure the charges, masses and magnetic moments of the atoms of the chemical elements; to study the processes of detaching atoms of electricity from atoms of matter; and to extend the spectra of the elements by detecting a host of radiations previously unknown and determining their frequencies. The data so assembled, together with observations upon the encounters of electrified particles with atoms, illuminated the relations between the elements, and contributed to the design of an atom-model which has already inspired many discoveries. Among these the greatest was the discovery of the Stationary States, which replaced the early way of interpreting spectra by a new and strikingly fruitful procedure, and taught experimenters to seek after and find a multitude of new phenomena of the most varied interest and importance in almost every field of physics. To name only two of the fields thus enriched: the flow of electricity through gases, and the conditions for the excitation of radiation, have been clarified in most unexpected fashion since the recognition of the Stationary States.

The book "Contemporary Physics," by Dr. Darrow, is devoted to these fundamental discoveries and to some of their consequences. It might be described as an introduction to the Theory of Atomic Structure, in the present day acceptance of this phrase; for the phenomena with which it deals have nearly all been used in designing atom models, and reciprocally a great many of them have been discovered in the course of making experimental tests of predictions based upon atom-models. These models are in fact among the most important features of contemporary physics and it would neither be possible nor desirable to omit them from such a book; for they are undoubtedly valuable, and the phenomena could hardly be described briefly and clearly without making use of them. Nevertheless, the actual facts of experience receive the greater emphasis, for they are the permanent and unassailable parts of the recent extensions of physics.

The book is developed from articles which have been appearing in the BELL SYSTEM TECHNICAL JOURNAL under the general title *Some Contemporary Advances in Physics*, apart from the articles concerning Hertzian waves and conduction of electricity in solids, which fall outside of the field to which it is restricted. The remaining articles, which were originally self-contained and separate, were largely rewritten in order to build out of them a coherent presentation of a unified field, and in the course of the rewriting they were nearly doubled in length by addition of new material. The book may be described as an elementary treatise, and although hardly of popular nature it will be intelligible to anyone who has had a fairly thorough college course in physics.

Published by D. Van Nostrand & Co. ($6.00).

Electric Circuit Theory and the Operational Calculus. JOHN R. CARSON. This book is based upon a course of fifteen lectures delivered recently at the Moore School of Electrical Engineering of the University of Pennsylvania.

The name of Oliver Heaviside is known to engineers the world over. His Operational Calculus, however, is known to and employed by only a relatively few specialists and this, notwithstanding its remarkable properties and wide applicability not only to electric circuit theory but also to the differential equations of mathematical physics. The present author ventures the suggestion that this neglect is due less to the intrinsic difficulties of the subject than to unfortunate obscurities in Heaviside's own exposition. In the present work, the Operational Calculus is made to depend on an integral equation from which the Heaviside rules and formulas are simply but rigorously deducible. It is his hope that this method of approach and exposition will secure a wider use of the Operational Calculus by engineers and physicists and a fuller and more just appreciation of its unique advantages.

The second part of the volume deals with advanced problems of electric circuit theory and in particular with the theory of the propagation of current and voltage in electrical transmission systems. It is hoped that this part will be of interest to electrical engineers generally because, while only a few of the results are original with the present work, most of the transmission theory dealt with is to be found only in scattered memoirs and there accompanied by formidable mathematical difficulties. While the method of solution employed in the second part is largely that of the Operational Calculus, the author has not hesitated to employ developments and explanations not to be found in Heaviside. For example, the formulation of the problem as a Poisson integral equation is an original development which has proved quite useful in the numerical solution of complicated problems. The same may be said of the chapter on Variable Electric Circuit Theory.

In view of its two-fold aspect, this work may therefore be regarded either as an exposition and development of the operational calculus with applications to electric circuit theory or as a contribution to advanced electric circuit theory depending upon whether the reader's viewpoint is that of the mathematician or the engineer.

No effort has been spared by the author to make his treatment clear and as simple as the subject matter will permit. The method of presentation is distinctly pedagogic. To electrical engineers and to electrical instructors this exposition of the fundamentals of electric circuit theory and the operational calculus should be of more than ordinary value. An appendix furnishes a list of original papers and memoirs which gives a fairly complete bibliographic survey of the field.

The volume is published by McGraw-Hill Book Co., price $3.00.

Exploring Life: the Autobiography of Thomas A. Watson. To have been the youth who at the age of twenty was assigned to build Alexander Graham Bell's original telephone apparatus, and then to share with him and Sanders and Hubbard the cares of rearing the telephone industry in the United States to healthy childhood, and finally to share the handsome profits which accrued therefrom, would doubtless satisfy the desires of the average ambitious individual. But to look upon this as merely a beginning and before the age of thirty to separate himself voluntarily from the business he had helped to found and set forth in quest of achievement in other and entirely unrelated fields attests an eagerness to play the game of life which cannot fail to be an inspiration to everyone.

So interesting is the life the author has led and so charmingly has he related his varied activities that the book would be welcome at any time, but coming during the semicentennial of the invention of the telephone, it is appropriate as well. To those who are desirous of obtaining further light on the early career of the telephone, particularly in the United States, the book brings several chapters of new material. But to the much wider circle who find enjoyment in a document which is at once homely and adventurous, every page of this autobiography will yield delight.

Published by D. Appleton & Co., price $3.50.

Some Measurements of Short Wave Transmission.[1] R. A. HEISING and J. C. SCHELLENG and G. C. SOUTHWORTH. Quantitative data on field strength and telephonic intelligibility are given for radio transmission at frequencies between 2.7 megacycles (111 m.) and 18 megacycles (16 m.) and for distances up to 1,000 miles, with some data at 3,400 miles. The data are presented in the form of curves and surfaces, the variables being time of day, frequency and distance. Comparisons are made between transmission over land and over water, between night effects and day effects, and between transmission from

[1] Proceedings of I. R. E., Oct., 1926.

horizontal and from vertical antennas. Fading, speech quality and noise are discussed. The results are briefly interpreted in terms of present day short wave theories.

An Introduction to Ultra-Violet Metallography.[2] FRANCIS F. LUCAS. This paper describes the ultra-violet microscope and the technique of its application to the study of metal surfaces. The ultra-violet microscope can be said to have lived up to expectations. Crisp brilliant images can be obtained which surpass in quality those obtainable with the apochromatic system. The potential resolving ability of the monochromats can be realized in practice and the practical application of the ultra-violet microscope should develop much new information. The ultra-violet microscope is the most complicated of any within the realm of technical or scientific microscopy. It requires a highly developed technique for its successful manipulation and the specimens must be prepared with great care. The ultra-violet equipment appears to have a potential resolving ability probably greater than twice that of the apochromatic system.

Portable Receiving Sets for Measuring Field Strengths.[3] AXEL G. JENSEN. Describes a measuring set involving the use of a current-dividing potentiometer accurate for frequencies up to about 1,500 kilocycles and having a field-strength range of about 20 to 200,000 microvolts per meter.

Thermionic and Adsorption Characteristics of Caesium on Tungsten and Oxidized Tungsten.[4] JOSEPH A. BECKER. Curves showing the logarithm of the electron current per cm^2 from tungsten and oxidized tungsten over a wide range of filament temperatures are given for several vapor pressures of caesium. At high temperatures the tungsten is covered only to a slight extent with adsorbed caesium. As the filament temperature is lowered more caesium is adsorbed. This lowers the electron work function and increases the emission many thousandfold. The process continues until a temperature is reached at which the tungsten is just covered with a monatomic layer when the work function has a minimum value. At still lower temperatures the surface is more than completely covered, the work function increases again, and the emission decreases rapidly.

[2] Presented before the American Institute of Mining and Metallurgical Engineers, New York, N. Y., February, 1926. Published as Pamphlet No. 1576-E, issued with *Mining and Metallurgy*.

[3] Proceedings I. R. E., page 333, June, 1926.

[4] *Physical Review*, Vol. 28, pp. 341–361, August, 1926.

The positive ion emission is constant while the temperature decreases from a high value to a low critical temperature. Here the ion emission drops suddenly while some caesium sticks to the filament. Further decreases in temperature are followed by increased adsorption and decreased ion emission. If the temperature is then increased in steps the ion current retraces its path. At an upper critical temperature, about 50° higher than the lower critical temperature, the filament cleans itself spontaneously, the caesium comes off as ions and registers as a sudden rush of current. At higher temperatures the ion current has its initial constant value which is limited by the arrival rate of caesium atoms. The critical temperatures are raised by an increase in the vapor pressure or by a decrease in the plate potential.

A method of determining the amount of adsorbed caesium is developed. At a sufficiently high filament temperature the surface is clean. At a sufficiently low temperature every atom that strikes the filament sticks to it, at least until the optimum activity is reached. The product of the arrival rate, which is given by the steady positive ion current, and the time to attain the optimum activity gives the number of caesium atoms at the optimum activity. At an intermediate temperature the surface is only partly covered. If the temperature is suddenly dropped, to a low value, it takes a shorter time to reach the optimum activity. From these times the amount of adsorbed caesium at various temperatures, plate potentials, and vapor pressures can be determined. At the optimum activity there are 3.7×10^{14} atoms of caesium on a cm^2 of tungsten. This is very nearly the same as the number of caesium atoms that could be packed in a single layer, but is considerably smaller than the number of caesium ions in such a layer.

The adsorption or evaporation characteristics are illustrated by curves. Caesium can evaporate either as ions or as atoms. The atomic rate depends only on the temperature and on Θ, the fraction of the surface covered with caesium. For a given temperature it increases very rapidly with Θ. The ions can permanently escape from the filament only if the potential is in the right direction. A typical isothermal ion rate curve increases rapidly with Θ, comes to a maximum when Θ is about .01, then decreases continuously for larger Θ. These curves explain all the observed phenomena of these adsorbed films. They show that while the ion work function increases as Θ increases, the work to remove an atom decreases with Θ. The ion work function for a given Θ can be decreased by increasing the potential gradient at the filament.

The Significance of Magnetostriction in Permalloy.[5] L. W. McKeehan. Magnetostriction in permalloy confirms qualitatively the existence of atomic magnetostriction as previously proposed, and the explanation, based thereon, for high magnetic permeability and low hysteresis in these alloys. The effect of tension upon magnetostriction suggests that orientation of the magnetic axes of iron and nickel atoms precisely like that due to the application of magnetic fields may be effected by mechanical stresses within the elastic limit. Acceptance of this view makes it possible to explain the large effects of tension upon magnetic hysteresis and the observed relation between the changes in electrical resistance produced by tension and by magnetization. The occurrence of a reversal of magnetostriction in a stretched wire containing 80 per cent nickel is covered by the same explanation. A connection between magnetic hysteresis and mechanical hysteresis is suggested and the molecular field postulated by Weiss is interpreted as the integrated effect of local mechanical stresses.

Magnetostriction in Permalloy.[6] L. W. McKeehan and P. P. Cioffi. The materials studied comprised iron, nickel, and permalloys containing 46, 64, 74, 78, 80, 84, and 89 per cent nickel. The method permitted simultaneous measurement of magnetization and magnetostriction in about 12 cm. at the middle of a 40 cm. wire, 1 mm. in diaeter, in an approximately uniform field not exceeding 40 gauss, and either with or without applied tension (within the elastic limit).

The magnetostriction was measured by a combination of a mechanical lever, an optical lever, a multiple slit and a photoelectric cell. The magnifying power of this combination, as used, was about 2×10^{6}, and magnetostrictive strains from 2×10^{9} to 3×10^{5} were detected and measured without changing the sensitivity.

The magnetostriction-magnetization curve has initial slope zero in all the cases studied. When the attainable field was sufficient for magnetic saturation the magnetostriction reached a limiting value. In iron there is evidence for the existence of a Villari reversal in fields too great to be attained in the apparatus. In nickel there is no sign of such reversal. In the permalloys with more than 81 per cent Ni the magnetostriction is a contraction. In the permalloys with less than 81 per cent Ni the magnetostriction is an expansion. The limiting values of megnetostriction, when plotted against chemical composition, fall on a smooth curve.

[5] *Physical Review*, Vol. 28, page 158, July, 1926.

[6] *Physical Review*, Vol. 28, page 146, July, 1926.

Tension increases magnetostrictive contraction and diminishes magnetostrictive expansion. It causes a reversal in the sign of magnetostriction in permalloy with 80 per cent Ni, a small contraction preceding the final small expansion.

Transmission Features of Transcontinental Telephony.[7] H. H. NANCE and O. B. JACOBS. In this paper, the various steps in the establishment of the existing network of transcontinental type circuits and the transmission design considerations are reviewed. The discussion covers the communication channels obtained from transcontinental type facilities and the bands of frequencies employed, and includes carrier-current systems, telephone repeaters and signalling systems. Mention is made of special uses of transcontinental telephone circuits, such as the transmission of program material for broadcasting and the transmission of pictures. Finally, the maintenance methods required to keep the system at full efficiency are outlined.

[7] Presented at the Pacific Coast Convention of the A. I. E. E., Salt Lake City, Utah, Sept. 6–9, 1926.

Contributors to this Issue

S. E. ANDERSON, B.S. in E. E., University of Michigan, 1919; Electrical Research Laboratory, General Motors Corporation, Detroit, 1919–20; Engineering Department, Western Electric Company, 1920–24; Bell Telephone Laboratories, 1925—. Mr. Anderson has been engaged in the development of carrier systems and radio receivers.

JOHN R. CARSON, B.S., Princeton, 1907; E.E., 1909; M.S., 1912; Research Department, Westinghouse Electric and Manufacturing Company, 1910–12; instructor in 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 many papers on theory of electric circuits and electric wave propagation.

E. T. HOCH, B.S., in Electrical Engineering, Case School of Applied Science, 1914; Western Electric Company, Manufacturing and Installation Department, 1914–15; Engineering Department, 1915–24; Bell Telephone Laboratories, Inc., 1925—.

F. C. WILLIS, B.Sc., London University, 1911; Engineering Department of the Western Electric Company, 1913–14; 1919–24; British Army, 1914–19; Bell Telephone Laboratories, 1925—. Mr. Willis was formerly engaged on design of radio apparatus and now on line and transmission testing apparatus.

L. E. MELHUISH, B.Sc., Pennsylvania State College, 1919; Engineering Department of the Western Electric Company and Bell Telephone Laboratories, 1919—. Mr. Melhuish has been employed on apparatus design relating to telephone repeaters, loud speaking systems and associated apparatus.

WALTER A. SHEWHART, A.B., University of Illinois, 1913; A.M., 1914; Ph.D., University of California, 1917; Engineering Department, Western Electric Company, 1918–24; Bell Telephone Laboratories, Inc., 1925—. Mr. Shewhart is making a special study of the application of probability theories to inspection engineering.

Frances Thorndike, A.B., Vassar College, 1922; M.A., Columbia University, 1925, American Telephone and Telegraph Company, Department of Development and Research, 1922—.

S. D. Wilburn—Long Lines Department of the American Telephone and Telegraph Company, 1904–1926; Line Construction, 1904–1906; Line Maintenance, 1906–1909; Equipment Maintenance, 1909–1912; Special Contract Service and Plant Maintenance, 1912–1918; Telephone Transmission Maintenance, 1918–1922; Telegraph Engineering, 1922–1926.

H. W. Hitchcock, B.S., Pomona College, 1911; Cornell University, 1911–13; Engineering Department of American Telephone and Telegraph Company, 1913–19; Department of Development and Research, 1919–21; The Pacific Telephone and Telegraph Company, Engineering Department, 1921–24; Transmission and Protection Engineer, Southern California Telephone Company, 1924—. Mr. Hitchcock's work prior to 1921 was in connection with inductive interference and submarine cable transmission development problems; his later work has been in connection with the transmission activities of one of the operating companies.