THE BELL SYSTEM

TECHNICAL JOURNAL

VOLUME XXXI JANUARY 1952 NUMBER 1

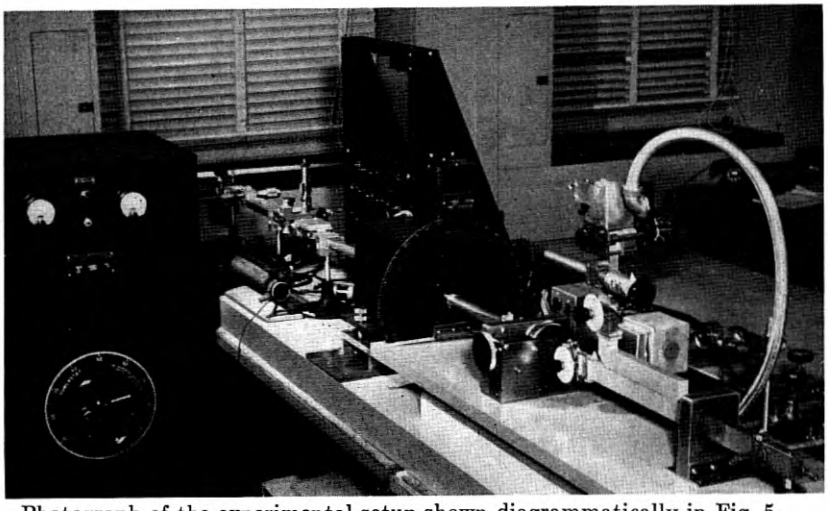
The Ferromagnetic Faraday Effect at Microwave Frequencies and its Applications

The Microwave Gyrator

BY C. L. HOGAN

A new microwave circuit element dependent on the Faraday rotation of a polarized wave has been developed. The element violates the reciprocity theorem and, because it shares this property with a gyroscope and because it is dependent on gyromagnetic resonance absorption, it has been termed a microwave gyrator. It is a low-loss broadband device with many applications. Among these are one-way transmission systems, microwave circulators, microwave switches, electrically controlled variable attenuators and modulators.

The microwave gyrator has been realized by making use of the Faraday rotation in pieces of ferrite placed in the waveguide. Polder has previously shown, in his analysis of the gyromagnetic resonance phenomenon, that ferromagnetic substances should show appreciable Faraday rotations at microwave frequencies. In the present study, Polder's analysis has been extended to include a wave being propagated through a ferromagnetic substance with dielectric and magnetic loss, and data are presented which give experimental verification of the theory. In addition an experimental technique is described which may be of some interest in studying the properties of ferrites at microwave frequencies.



Photograph of the experimental setup shown diagrammatically in Fig. 5.

INTRODUCTION

In a recent series of articles, Tellegen[1] has discussed the possible applications of a new circuit element which he calls a gyrator. He defines the ideal gyrator, in principle, as a passive four-pole element which is described by: (see Fig. 1)

$$v_1 = -Si_2 \qquad v_2 = Si_1 \tag{1}$$

Since the coefficients above are of opposite sign, the gyrator violates the theorem of reciprocity. Any network composed of the usual electrical circuit elements—resistors, inductors, capacitors, and transformers—will satisfy the theorem of reciprocity. In simple terms, this theorem states that if one inserts a voltage at one point in the network and measures the current at some other point, their ratio (called the transfer impedance) will be the same if the positions of voltage and current are interchanged. In the gyrator, however, this transfer impedance for one

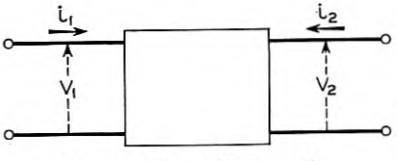


Fig. 1—General four-pole.

direction of propagation is the negative of that for the other direction of propagation. Essentially this means that a 180° phase difference exists between the two directions of propagation. For this reason it has been suggested that the element could be more aptly called a directional inverter.[2]

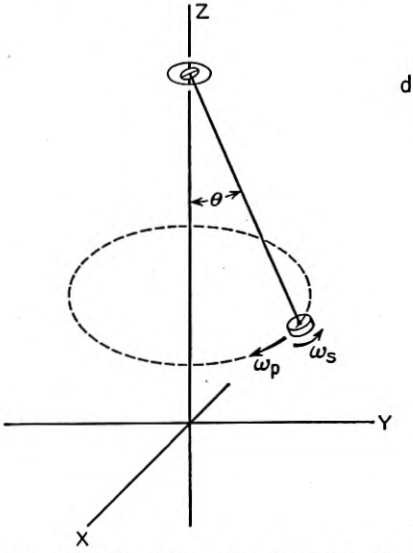
Network synthesis today is based upon the existence of four basic circuit elements: the capacitor, the resistor, the inductor, and the ideal transformer. It is apparent that the introduction of a fifth circuit element, the gyrator, would lead to considerably improved solutions for many network problems. In fact, Tellegen[1] has shown that the synthesis of resistanceless four-pole networks would be much simplified by its introduction. In addition, McMillan[3] has shown that it would be possible to construct a one-way transmission system if a gyrator were available, and Miles[4] has shown that it would be possible by use of a gyrator to construct a network which is equivalent to a Class A vacuum tube amplifier circuit. While the realizable power gain of these gyrator circuits is necessarily always not greater than unity, many other networks including gyrators are possible which have properties analogous to vacuum tube circuits and some of these may be of practical importance. Since this new element offers such interesting possibilities in network synthesis, a study has recently been made in these Laboratories of possible methods for realizing the gyrator.

A gyrator was employed by Bloch[5] in his measurement of the magnetic moment of the proton. Bloch made use of the phenomenon that if two crossed coils with a mutual core are adjusted so that there is zero mutual inductance between them and if a steady magnetic field is applied perpendicular to the axes of both coils, then an ac voltage applied to one of the coils will induce a voltage in the second due to the gyromagnetic resonance phenomenon. This induction is ordinarily extremely small unless the magnetic field is adjusted so that the exciting frequency coincides with a gyromagnetic resonance frequency of the material which forms the mutual core of the two coils. In Bloch's experiment, the magnetic field was held constant and the exciting frequency was adjusted until it coincided with the gyromagnetic resonance frequency of the proton. If they were wound over a paramagnetic or ferromagnetic material, the two crossed coils would form a gyrator when the magnetic field was adjusted so that the frequency of the exciting field coincided with the gyromagnetic resonance of the unpaired electrons. The fact that this structure constituted a gyrator was first recognized by Tellegen[1] and has been discussed by Beljers and Snoek[6]

in a paper which gives a very satisfying physical model with which to interpret gyromagnetic phenomena occurring within ferrites.

Physical analysis indicates that the properties of ferromagnetic materials can be explained by assuming that the electron behaves as if it were a negatively charged sphere which is spinning about its own axis with a fixed angular momentum. This rotation of charge imparts to the electron a magnetic moment which is a function of the electric charge on the electron, the angular velocity of the electron, and its size. Thus the electron behaves as if it were a spinning magnetic top, whose magnetic moment lies along the axis of rotation, and its behavior can be understood by considering a spinning gyroscope suspended in gimbal rings at a point not coinciding with its center of gravity. If a gyroscope, thus supported in a gravitational field, is lifted away from its position of minimum potential energy and then released, it will not return to the position of minimum energy but will precess about the vertical axis. This is illustrated in Fig. 2 where the spinning gyroscope makes an angle θ with the vertical z, axis. Its equilibrium motion, in the absence of damping, is a precessional motion about the vertical axis with a velocity ω_p.

If the gyroscope be regarded as initially hanging vertically downward



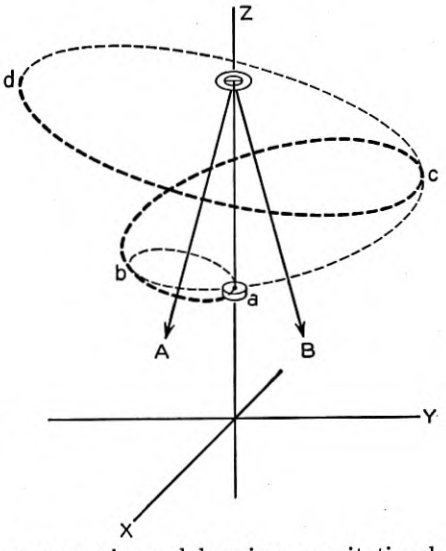


Fig. 2 (left)—Precessional motion of a gyroscopic pendulum in a gravitational field. Fig. 3 (right)—Precessional motion of a gyroscopic pendulum in a gravitational field which oscillates between the directions A and B.

as indicated in Fig. 3 and then a gravitational force is suddenly made to act along the y axis so that the net gravitational force acts along A, it is obvious that the gyroscope will begin to precess about the gravitational field direction as indicated by the small dotted circle. However, if after completing a half cycle, the horizontal component of the gravitational field is reversed so that now the net gravitational field acts along the vector B, the gyroscope will begin to precess about B as indicated by the intermediate size dotted circle. If the horizontal component of the gravitational field is again reversed after the gyroscope completes another half cycle in its precession, the gyroscope will again begin to precess about the direction A and the actual path of the precessional motion will be along the path *a-b-c-d*. If this process is continued indefinitely, the gyroscope will precess in larger and larger circles around the vertical until the damping becomes large enough to contain the gyroscope in some equilibrium circle (assuming that the damping is large enough to accomplish this).

The above model affords a classical picture which can be used quite readily to describe the motion of the electrons in a ferrite. If the ferrite is initially saturated along the z axis by a steady magnetic field, the electrons will come to rest with their magnetic moments lying along the z axis, as the gyroscope in Fig. 3. If now an alternating magnetic field is applied along the y axis, the electrons will begin to precess in larger and larger circles about the z axis until they finally reach some equilibrium position under the influence of the magnetic fields and the damping. Thus it is apparent in the gyromagnetic resonance experiments described above why an alternating field applied perpendicular to a steady magnetic field in a ferrite will give rise to a varying flux perpendicular to both the steady field and the alternating field. It is also apparent why the alternating flux along the x axis is 90° out of phase with the alternating flux along the y axis. Since precession of the top will always be in the same direction regardless of whether the alternating field is applied along the x or y axes, consideration of Fig. 3 makes it apparent how the two crossed coils with ferrite at their center can constitute a gyrator which violates the reciprocity relation in a manner described by Equations (1). To the present time, however, no practical circuit element making use of this phenomenon has been constructed because the coefficient of coupling between the coils is always small, even in the vicinity of the resonant frequency, and also because the losses in the materials available are so high in the vicinity

of resonance that the insertion loss of such a device would be prohibitively large.

McMillan in his original article,[3] showed that a gyrator could be realized by means of mechanically coupled piezo-electric and electromagnetic transducers. Later, McMillan[7] pointed out that a gyrator could be realized by means of the Hall effect in a square plate of bismuth, as was also predicted by Casimir.[7] Another similar possibility would be an electrical-electrical coupling through a gyroscopic link. A gyrator has been built by W. P. Mason of these Laboratories which makes use of the Hall effect in a crystal of germanium.[8] This gyrator showed an insertion loss somewhat higher than the theoretical loss of 12.3 db. R. O. Grisdale of these Laboratories suggested that these losses could be greatly reduced if the same Hall effect principle were applied to a vacuum tube which contained four electrodes which could both emit and collect electrons. This device is no longer passive, but such a structure has been built and showed an insertion loss of about 7 db, only slightly higher than the theoretical loss which would be expected from this geometry.

In view of the substantial losses found to exist in the earlier forms of gyrator discussed above, a study of other "anti-reciprocal" phenomena which might lead to the realization of a relatively low loss gyrator was undertaken.

It has long been known that the Faraday rotation of the plane of polarization in optics is anti-reciprocal. In order to observe the Faraday rotation, polarized electromagnetic waves must be transmitted through a transparent isotropic medium parallel to the direction of the lines of force of a magnetic field. The effect is usually produced by placing the material along the axis of a solenoid. The rotation is "positive" if it is in the direction of the positive electric current which produces the field and "negative" if in the opposite direction. All optically transparent substances show the Faraday rotation.

Its anti-reciprocal property distinguishes the Faraday effect from optical rotations caused by birefringent crystals, or by the Cotton-Mouton effect, which are reciprocal. That is, if a plane polarized light-wave is incident upon a birefringent crystal in such a manner that the plane of polarization is rotated through an angle θ in passing through the crystal, then this rotation will be cancelled if the wave is reflected back through the crystal to its source. In the Faraday rotation, however, the angle of rotation is doubled if the wave is reflected back along its path. Hence, if the length of path through the "active" material is adjusted so as to give a 90° original rotation, the beam on being reflected

will have its plane of polarization rotated a total of 180° in passing in both directions through the material. Thus, the Faraday rotation in optics affords an anti-reciprocal relation quite analogous to the anti-reciprocal property of the gyrator postulated by Tellegen.

Lord Rayleigh[9] described a one-way transmission system in optics which makes use of the Faraday rotation. Lord Rayleigh's "one-way" system consisted of two polarizing Nicol prisms (oriented so that their planes of acceptance made an angle of 45° with each other), with the material causing the Faraday rotation placed between them. Thus, light which was passed by the first crystal and whose plane of polarization was rotated 45° would be passed by the second crystal also. But, in the reverse direction, the rotation would be in such a sense that light which was admitted to the system by the second crystal would not be passed by the first.

Although Rayleigh's one-way transmission system can be actually realized, it is experimentally difficult since most substances show extremely small Faraday rotations. In fact, large rotations for transparent substances in the optical region are of the order of one degree per cm path length for an applied magnetic field of 1000 oersteds. To realize a rotation of 45° would require maintaining a field of 1000 oersteds over a distance of approximately one-half meter. The Faraday effect in ferromagnetic substances, however, is unique in that it shows rotations many orders of magnitude greater than the rotations exhibited by any other substances. For instance, König[10] reports rotations of 382,000°/cm by passing light through thin layers of magnetized iron. These data, of necessity, however, were taken on extremely thin sections and the total rotation obtained for any specimen did not exceed 10°. In order to obtain appreciable rotations in a device of practical size, it is necessary to obtain a material which shows a rotation at least intermediate between those reported for iron and other ordinary materials. In addition, in order to make effective use of these rotations, the material must be transparent to the radiation which is being used.

THEORY OF THE FERROMAGNETIC FARADAY EFFECT

Polder[11] has shown in his analysis of the ferromagnetic resonance phenomenon, that a plane electromagnetic wave at microwave frequencies should show appreciable Faraday rotation when propagated through a ferromagnetic material which is magnetized in a direction parallel to the direction of propagation of the wave. Polder has neglected both magnetic and dielectric losses in his analysis and although for the ferrites which are of greatest interest, this approximation is quite

valid, nevertheless the more complete theory is developed below. The exact theory of this phenomenon should, of course, be approached through quantum mechanics, but since the classical theory, in this particular case, gives a result as satisfactory as the quantum theory and since it lends itself more aptly to a fundamental physical interpretation of the phenomenon, it is the classical theory which is developed here. Quantum mechanically, Faraday rotations in the optical region are accounted for by the Zeeman splitting of the spectral lines.

The classical model which proves quite adequate for the description of ferromagnetic resonance is that illustrated in Figs. 2 and 3 which regards the electrons of the material which contribute to the magnetism as being spinning magnetic tops. The angular momentum of each electron is:

$$|J| = \tfrac{1}{2}(h/2\pi) \tag{2}$$

$\vec{J}$ = Angular momentum of electron (gm cm^2/sec)

h = Planck's constant (6.62×10^{-27} erg sec)

The magnetic moment which arises due to this rotation is:

$$|\mu_B| = \frac{eh}{4\pi mc} \tag{3}$$

where:

$\vec{\mu}_B$ = Magnetic moment of electron (Bohr magneton)

e = Charge on electron (4.80×10^{-10} E.S.U.)

m = Mass of electron (9.10×10^{-28} gm)

c = Velocity of light (3×10^{10} cm/sec)

The so-called gyromagnetic ratio of the electron is the ratio of these quantities and is given by:

$$\gamma = 2\,\frac{e}{2mc} = \frac{|\mu_B|}{|J|} \tag{4}$$

If a steady magnetic field is applied to the sample such that the electron sees an effective field H, then a torque will be applied to the electron which tries to turn the electron so that its magnetic moment lies along the field direction. However, as indicated in Fig. 2, the electron will precess around the field direction until damping forces dissipate the energy of precession. The equation of motion of the electron is:

$$\vec{\mu}_B \times \vec{H} = \frac{d\vec{J}}{dt} = \gamma^{-1}\frac{d\vec{\mu}_B}{dt} \tag{5}$$

The equation of motion of the magnetization per unit volume can thus be written:

$$\frac{d\vec{M}}{dt} = \gamma \vec{M} \times \vec{H} \tag{6}$$

where:

$\vec{M}$ = Magnetization of medium

$\vec{H}$ = Macroscopic internal magnetic field

The above equation, however, does not include damping. The damping force, regardless of its origin, must be so introduced into the above equation that it tends to cause the electron's axis of rotation to line up with the field direction. It has been shown by Yager, Galt, Merritt and Wood[12] that the shape of the resonance absorption line can be accounted for if the damping term is introduced in the following way:

$$\frac{d\vec{M}}{dt} = \gamma \vec{M} \times \vec{H} - \frac{\gamma \alpha}{|M|} [\vec{M} \times (\vec{M} \times \vec{H})] \tag{7}$$

The vector $\vec{M} \times (\vec{M} \times \vec{H})$ is simply a vector which is in the proper direction to act as a damping force (torque) and the coefficient is chosen so as to give the correct units along with the parameter, α, which must be determined experimentally and which gives the magnitude of the damping torque.

Equation 7 then is the equation of motion of the magnetization of an arbitrarily shaped body under the action of an arbitrary internal field, H. In the appendix, it is shown that if a steady magnetic field, H_a, is applied along the z axis and then a small alternating field is applied in an arbitrary direction to a sample which is infinite in size, the equation relating the resulting alternating flux density, b, and the applied alternating field, h is:

$$\begin{aligned} b_x &= \mu h_x - jKh_y \\ b_y &= jKh_x + \mu h_y \\ b_z &= h_z \end{aligned} \tag{8}$$

where

$$\mu = \mu' - j\mu'' \tag{9}$$

$$K = K' - jK'' \tag{10}$$

Equations which give μ and K in terms of the applied magnetic field and fundamental atomic constants are given in the appendix.

Equations (8) are easily interpreted in terms of the spinning gyroscope model of Fig. 3. If magnetic losses had been ignored (i.e. $\alpha = 0$) then both μ and K would have been real. Under this condition, it is seen that if an alternating field, h_y, is applied along the y axis, then an alternating flux, b_y, is created along the y axis which is in phase with h_y, and an alternating flux, b_x, is created which is 90° out of phase with h_y. Reciprocity between the x and y directions would demand that both terms containing jK should have the same sign. Thus, Equations (8) give a quantitative expression for the results which were previously qualitatively deduced by means of the electronic model illustrated in Figs. 2 and 3.

If a waveguide is filled with a ferromagnetic material such as a ferrite and if then a steady magnetic field is applied along the axis of the waveguide, it is necessary in order to describe this wave to find a solution to Maxwell's equations which is consistent with Equations (8) and in which b, h, E and D are all proportional to $\exp [j\omega t - \Gamma z]$. This problem is not solved exactly. However, in the appendix a solution is obtained for an infinite plane wave. It is found that the ferromagnetic medium can support only a positive or a negative* circularly polarized wave or a combination of both. It is also shown in the appendix that the propagation constants for these two circularly polarized waves are different and are given by the following expressions:

$$\Gamma_+ = \frac{j\omega}{c}\sqrt{(\mu + K)[\varepsilon]} \tag{11}$$

and

$$\Gamma_- = \frac{j\omega}{c}\sqrt{(\mu - K)[\varepsilon]} \tag{12}$$

where

$\Gamma_\pm$ = Propagation constant

ω = Angular frequency of wave

c = Velocity of light in unbounded space (3×10^{10} cm/sec)

ε = Complex dielectric constant of medium

In Equations (11) and (12) it is apparent that the effective permeability of the medium to a positive circularly polarized wave, for in-

* The usual notation is used here, where the positive component is the component which is rotating in the direction of the positive electric current which creates the steady longitudinal field.

stance, is given by the expression $(\mu + K)$, and not by the usual permeability, $b_x/h_x = \mu$. It is also apparent that the quantity $\mu + K$ can vary over wide limits in the vicinity of the ferromagnetic resonance. For this reason, care must be taken in interpreting permeability data for ferromagnetic materials which now occur in the literature and which were obtained by means of impedance measurements at microwave frequencies, since the above equations indicate that this method does not measure the same quantity that is measured at low frequencies by means of a toroidal sample overwound with two coils. The low frequency measurement of permeability obviously measures the quantity which is designated as μ in Equation (8).

If Equations (11) and (12) are solved for the attenuation constants, $\alpha_\pm$, and the phase constants, $\beta_\pm$, the following results are obtained:

$$\alpha_\pm = \frac{\omega}{c}\sqrt{\frac{(\mu' \pm K')\varepsilon'}{2}} \cdot \left[\left\{\sqrt{(1 + \tan\delta_m[4\tan\delta_d + \tan\delta_m(1 + \tan^2\delta_d)] + \tan^2\delta_d}\right\} \cdot - 1 - \tan\delta_m\tan\delta_d\right]^{\frac{1}{2}} \tag{13}$$

and

$$\beta_\pm = \frac{\omega}{c}\sqrt{\frac{(\mu' \pm K')\varepsilon'}{2}} \cdot \left[\left\{\sqrt{(1 + \tan\delta_m[4\tan\delta_d + \tan\delta_m(1 + \tan^2\delta_d)[+ \tan^2\delta_d}\right\} \cdot + 1 + \tan\delta_m\tan\delta_d\right]^{\frac{1}{2}} \tag{14}$$

where:

$$\tan\delta_m = \frac{\mu'' \pm K''}{\mu' \pm K'}$$

(The + sign must be used for a positive circularly polarized wave; the negative sign for the negative circularly polarized wave.)

$$\tan\delta_d = \frac{\varepsilon''}{\varepsilon'} = \text{dielectric loss tangent}$$

$$\varepsilon = \varepsilon' - j\varepsilon'' = \text{complex dielectric constant}$$

It is almost impossible to get a feeling for what these equations mean with respect to a wave travelling through the medium, especially since μ and K are given by equations which are almost as difficult to perceive. An appreciation of these equations can be obtained however, by reference to Fig. 4 which gives qualitatively the behavior predicted by these expressions. Essentially, α and β are functions of two variables. These are ω, the frequency of the wave, and H_a, the applied magnetic field. In Fig. 4, the index of refraction and attenuation of the positive circularly polarized component are given relative to these values for

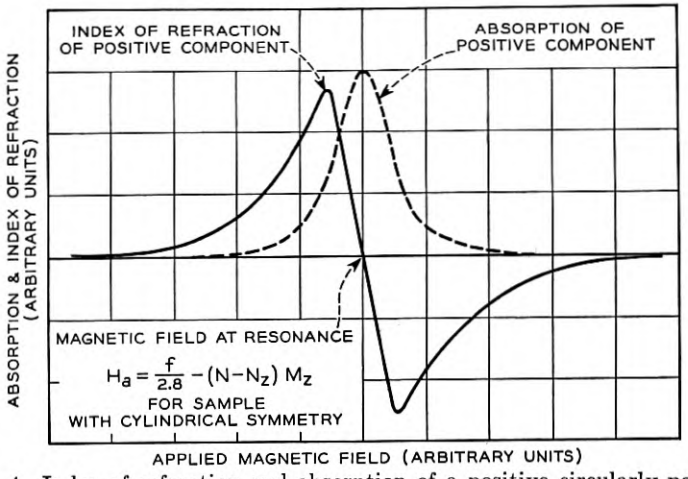


Fig. 4—Index of refraction and absorption of a positive circularly polarized wave relative to the same quantities for a negative circularly polarized wave being propagated through a magnetized medium.

the negative component. Hence, both the index of refraction and attenuation of the negative component are represented by the abscissa of the graph. In Fig. 4 these quantities are plotted as a function of the applied magnetic field for a wave of a fixed frequency. Many of the properties of the medium are clearly displayed in this graph. In particular, as the field necessary for ferromagnetic resonance is approached, the attenuation of the positive component becomes larger and larger. Eventually this component will be substantially completely absorbed and only the negative circularly polarized component will be propagated. Hence it should be possible to establish a circularly polarized wave in a waveguide simply by passing the dominant mode through a ferromagnetic material which is subjected to a longitudinal magnetic field of the proper amplitude. However, there will be an absorption of one-half of the power being propagated. If Fig. 4 had been plotted as a function

of the frequency of the wave for a fixed magnetic field, a similar set of curves would have resulted. This set would indicate the frequency dependence of the Faraday rotation. If the frequency of the wave is far removed from the resonance frequency, the difference between the indices of refraction of the positive and negative component is not frequency dependent. However, near resonance, this difference is a very rapidly varying function of the frequency. It is to be remembered that these equations were derived for an infinite plane wave. However, it would be expected that these equations would describe quite accurately the propagation of the dominant mode in a waveguide. The approximation would, of course, be better when the cut-off wavelength was much greater than the unbounded wavelength. This condition is met when the waveguide is filled with ferrite and for these cases quantitative agreement is obtained.

The above analysis shows that if a dominant mode wave (plane polarized) is incident upon a ferromagnetic material which is magnetized along the length of the waveguide, the wave will split into positive and negative circularly polarized waves whose phase constants are given by Equation (14).) Since the two circular components travel with different velocities in the medium, they will upon emerging from it unite to form a plane polarized wave whose plane of polarization has been rotated with respect to the incident polarization. The angle of rotation of the polarization is given by:

$$\theta = \frac{\ell}{2}[\beta_- - \beta_+] \tag{15}$$

where:

ℓ = path length through ferromagnetic material (cm)

In order to evaluate Equation (15), it must be combined with Equation (14). However, a few approximations are valid in Equation (14) which make it much simpler. In particular many ferrites exist for which the magnetic losses are extremely small as long as the internal field within the body is kept small so that the frequency of the wave does not approach the ferromagnetic resonance frequency. This field can be kept small if the magnetic field is not raised above the point necessary to saturate the ferrite. Kittel has shown that for a finite body the effective internal magnetic field that determines the resonant frequency is given by:

$$H_e^2 = [H_a + (N_x - N_z)M_z][H_a + (N_y - N_z)M_z]$$

where H_e is in oersteds. The ferromagnetic resonance frequency is given by:

$$f_{res} = 2.8H_e\sqrt{1 + \alpha^2} \text{ megacycles} \tag{16}$$

It is easily shown that the following formula is approximately valid for a ferromagnetic body with a circular or square cross-section

$$H_e = \frac{4\pi + N(\mu - 1)}{4\pi + N_z(\mu - 1)} H_a \tag{17}$$

at saturation. Where:

μ = true dc permeability at saturation

N = demagnetizing factor in x and y directions.

If an average value of 1000 is assumed for the dc permeability, then H_e can be readily computed for various shapes.

For a thin disc:

$$N = 0 \qquad N_z = 4\pi$$

and

$$H_e = \frac{H_a}{1000}$$

If a thin disc saturates at 1500 gauss, then:

$$H_e = 1.5 \text{ oersteds}$$

$$f_{res} \approx 4.2 \text{ megacycles}$$

For a long thin pencil:

$$N_z = 0 \qquad N = 4\pi$$

and

$$H_e = 1000H_a$$

For this case the body could be saturated with a field of about 1.5 oersteds, so:

$$H_e = 1500 \text{ oersteds}$$

and

$$f_{res} = 4200 \text{ megacycles}$$

If, for this case, the resonance frequency is so close to the operating frequency that losses due to ferromagnetic resonance become pro-

hibitive, it is wise to then raise the applied field to some high value, so that the resonance frequency will fall well above the operating frequency. Thus, for many cases of interest it is possible by various means to place the ferromagnetic resonance absorption frequency sufficiently far from the operating frequency so that magnetic losses due to this phenomenon are negligible.* The data accumulated to date indicate that the major component of the magnetic losses at microwave frequencies is due to this phenomenon. Only in a few cases have data been taken which have indicated that other factors, such as domain wall relaxation, contribute to the magnetic loss at microwave frequencies.

If then, the magnetic field is controlled so that the ferromagnetic resonance absorption is negligible, Equations (13) and (14) can be simplified to:

$$\alpha_{\pm} = \frac{\omega}{c}\sqrt{\frac{(\mu' \pm K')\varepsilon'}{2}}\quad\sqrt{\sqrt{1+\tan^2\delta_d}-1} \tag{18}$$

and:

$$\beta_{\pm} = \frac{\omega}{c}\sqrt{\frac{(\mu' \pm K')\varepsilon'}{2}}\quad\sqrt{\sqrt{1+\tan^2\delta_d}+1} \tag{19}$$

which can be written as:

$$\alpha_{\pm} = \frac{\omega}{c}\sqrt{\frac{|\varepsilon|-\varepsilon'}{2}}\quad\sqrt{\mu' \pm K'} \tag{20}$$

and

$$\beta_{\pm} = \frac{\omega}{c}\sqrt{\frac{|\varepsilon|+\varepsilon'}{2}}\quad\sqrt{\mu' \pm K'} \tag{21}$$

where μ' and K' are given in the appendix.

If Equation (21) is now inserted into Equation (15) a formula for rotation is obtained which is valid within the limits of the above approximations. If in addition, the frequency of the wave is sufficiently greater than the resonance frequency, so that:

$$\omega_{res} \ll \omega \tag{22}$$

then Equation (15) takes the particularly simple form:

$$\frac{\theta}{l} = \frac{\omega}{2c}\sqrt{\frac{|\varepsilon|+\varepsilon'}{2}}\left[\sqrt{1+\frac{4\pi M_z\gamma}{\omega}}-\sqrt{1-\frac{4\pi M_z\gamma}{\omega}}\right] \tag{23}$$

Most ferrites saturate at 2,000 gauss or less. Hence, for a frequency of

* This is not always possible, for some ferrites, in the polycrystalline state, exhibit extremely broad ferromagnetic resonance absorption lines and it is difficult to operate at any frequency without appreciable absorption.

9,000 megacycles,

$$\frac{4\pi M_z\gamma}{\omega} \leqq \frac{2000 \times 17.6 \times 10^6}{9000 \times 2\pi \times 10^6} = 0.622 \tag{24}$$

Hence, the following approximation will be valid to within 5 per cent.

$$\sqrt{1 \pm \frac{4\pi M_z\gamma}{\omega}} \approx 1 \pm \frac{1}{2}\left(\frac{4\pi M_z\gamma}{\omega}\right)$$

With this approximation, Equation (23) reduces to:

$$\frac{\theta}{\ell} = \frac{1}{2c}\sqrt{\frac{|\varepsilon| + \varepsilon'}{2}}\,[4\pi M_z\gamma] \tag{25}$$

Equation (25) is quite remarkable. Not only does it predict large rotations, but it also predicts that, within the above approximations the rotation will not depend upon the frequency of the incident radiation. For the assumed values,

$$\varepsilon' = 15$$

$$\varepsilon'' = 0$$

$$4\pi M_z = 1000,$$

Equation (25) predicts rotations of,

$$\frac{\theta}{\ell} = 65°/\text{cm}.$$

DESCRIPTION OF EQUIPMENT AND MEASURING TECHNIQUES

The Faraday rotation has been measured in a large number of ferrites in order to verify the above theory and in an effort to improve the characteristics of the microwave gyrator. A diagram of the experimental equipment is given in Fig. 5, and a diagram of the test chamber in which the rotations were measured is given in Fig. 6. In the test chamber, two rectangular waveguides are separated by a circular waveguide, the proper nonreflective transitions being made at each end of the circular section, which is about twelve inches long. One rectangular guide is supported so that it can be rotated about the longitudinal axis of the system. The dominant TE_{10} mode is excited in one rectangular guide, and by means of the smooth transition this goes over into the dominant TE_{11} mode in the circular guide. The rectangular guide on the opposite end will accept only that component of the polarization which coincides with the TE_{10} mode in that guide, the other component being reflected

at the transition. Absorbing vanes, inserted in the circular section, absorb this reflected component. The circular guide is placed in a solenoid to establish an axial magnetic field along its length.

The ferrite cylinders to be measured were placed at the mid-section of the circular guide. When a cylinder was used which did not fill the cross-section of the guide, it was supported along the axis of the guide by means of a hollow polystyrene cylinder which did fill the guide.

In addition to measuring the Faraday rotation, measurements of insertion loss were made by determining the power transmitted under identical conditions with the ferrite cylinder removed, and the ellipticity of the transmitted wave was determined by measuring the power transmitted when the rectangular guide on the detector side was rotated to both positions of maximum and minimum transmission. Power transmission measurements could be repeated within 0.2 db. Measurements of the angle of rotation of the plane of polarization could be repeated within $\frac{1}{2}°$ except in the region close to the gyromagnetic resonance where rotations were large and ellipticity so great that it was difficult to decide the positions of maximum and minimum transmission. These errors increased up to the point where the transmitted wave was circularly polarized where it was impossible to measure the angle of rotation.

EXPERIMENTAL RESULTS

Equation (25) indicates that the rotation per unit path length through the ferromagnetic material is proportional to the magnetization of the

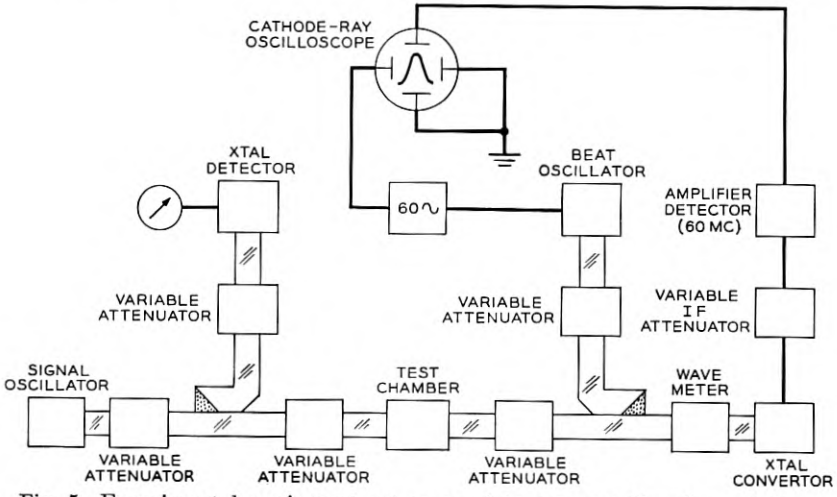


Fig. 5—Experimental equipment set-up used to measure Faraday rotations.

sample and is not dependent directly on the applied magnetic field. Fig. 7 shows the dependence of rotation upon magnetization for a sample of manganese zinc ferrite, and indicates that after the sample is saturated, the rotation is sensibly independent of the applied magnetic field. In addition, the complex dielectric constant and the saturation magnetization of this sample were measured. From these the rotation per centimeter path can be computed from the above theory using

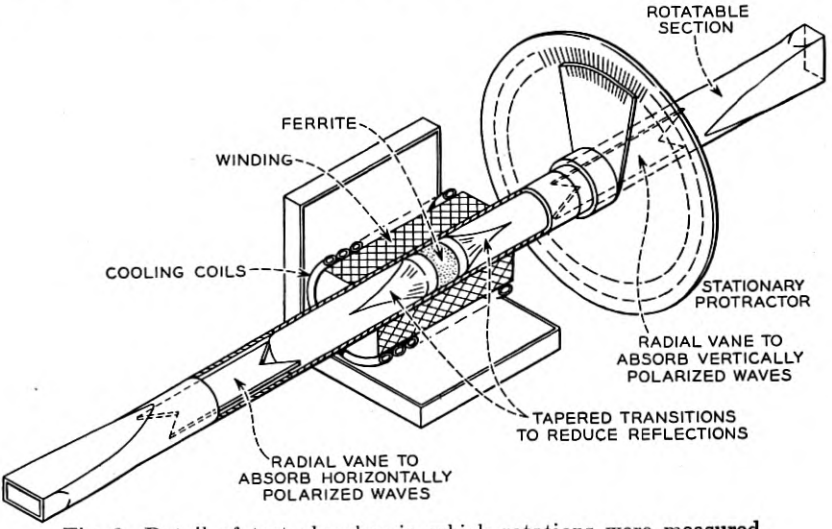


Fig. 6—Detail of test chamber in which rotations were measured.

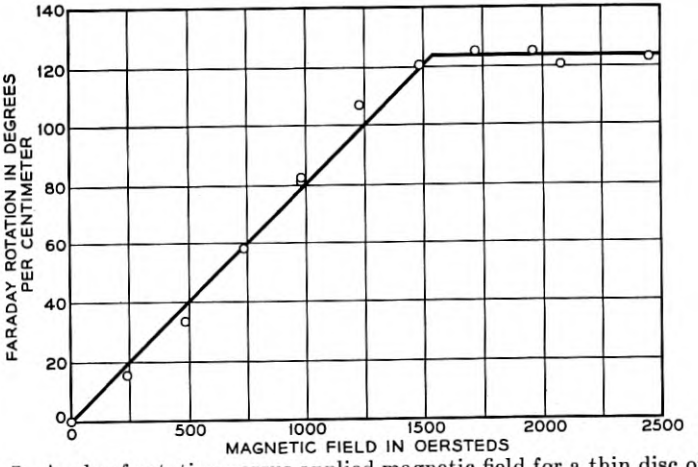


Fig. 7—Angle of rotation versus applied magnetic field for a thin disc of manganese zinc ferrite.

Equation (25). For a particular sample of manganese zinc ferrite, the following measurements were made:

(Sample No. 1)

$$\varepsilon' = 17$$

$$\varepsilon'' = 24$$

$$4\pi M_{sat} = 1500 \text{ gauss}$$

Using this data, equation (25) predicts:

$$\frac{\theta}{\ell} = 121.2°/\text{cm}.$$

It is seen in Fig. 6, that the actual measured rotation at saturation is approximately 123°/cm. Hence an extremely good agreement with theory has been obtained for this particular sample.

Equation (25) also indicates that the rotation per unit path length should be sensibly independent of frequency within the above approximations. The data are shown in Fig. 8. However, it will be noticed that the frequency difference between these two sets of data is relatively small (3 per cent), and the cumulative experimental error in measuring angles is such that it is difficult to state that the rotation is closer than 1° between the two sets of data. This represents a possible difference of 5 per cent in the rotation for a change of 3 per cent in the frequency. Thus, even though these preliminary data support Equation (25), it cannot be accepted as conclusive evidence until more measurements can be made over a wider band width.

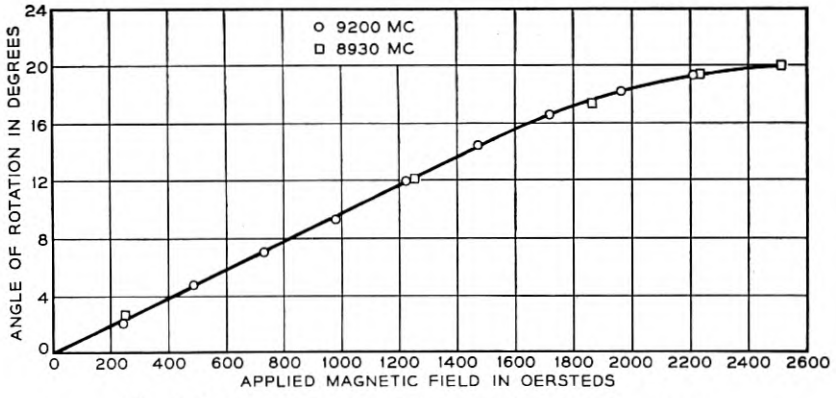


Fig. 8—Dependence of Faraday rotation upon frequency.

The loss characteristics of different ferrites as a function of the applied magnetic field differed distinctly from each other. Some ferrites, such as manganese zinc ferrite showed extremely high loss which was associated with the imaginary part of the dielectric constant. This loss was not affected by the application of a magnetic field but remained substantially constant as the field was applied. However, as the field approached that necessary for ferromagnetic resonance, the total power absorbed by the ferrite increased, since the positive circularly polarized component was almost completely absorbed by the sample. In fact by measuring the ellipticity of the transmitted wave, it is possible to compute the difference between the absorption of the positive and negative circularly polarized components. This has been done for Sample No. 1 and the result is indicated in Fig. 9. If the curve were continued to higher fields, it would represent the shape of the ferromagnetic resonance absorption line.

Some ferrites, such as Ferramic G, showed an almost zero dielectric loss but on the other hand caused an extremely large absorption at 9000 megacycles due to magnetic losses. The major contributions to magnetic loss at this frequency should be either losses associated with a domain wall relaxation or ferromagnetic resonance absorption due to anisotropy fields. Unequivocal data can be obtained by the above techniques to identify which loss is predominant. If the loss were due to domain wall relaxation (or resonance) it would absorb both the negative and positive circularly polarized components equally. Thus as the magnetic field was applied and as the ferrite became saturated, the losses in both components should decrease as the domain walls disappeared. However,

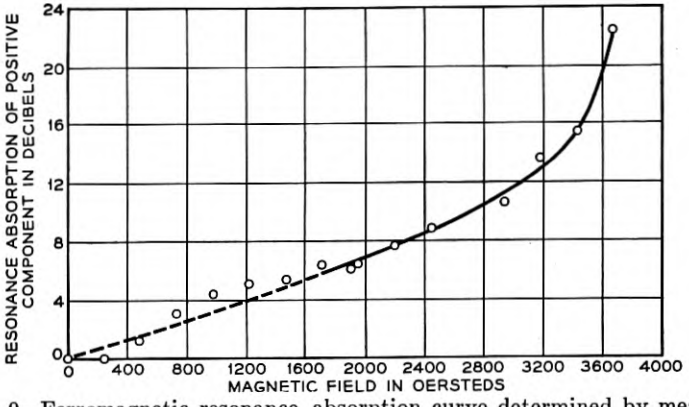


Fig. 9—Ferromagnetic resonance absorption curve determined by measuring the ellipticity of a wave transmitted through a cylinder of ferrite in a waveguide.

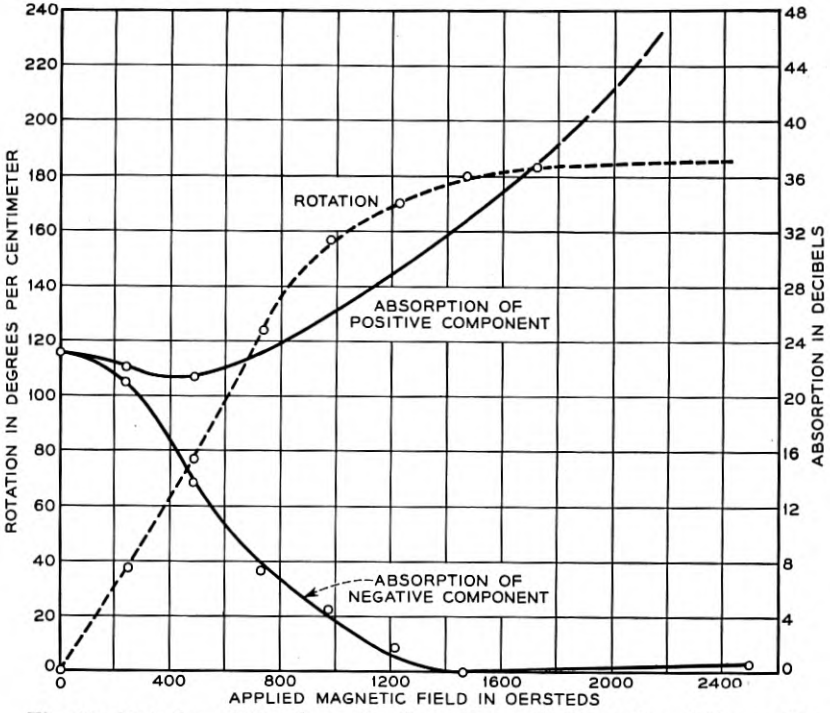


Fig. 10—Rotation of the plane of polarization and absorption of the positive and negative circularly polarized components when an electromagnetic wave (X-band) is propagated through a cylinder of Ferramic G.

if instead the loss is associated with ferromagnetic resonance absorption, the absorption of the positive component should begin to increase as soon as more domains are lined up in a direction where they can absorb the positive component. Thus even before the sample is saturated the absorption of the positive component should be much larger than the absorption of the negative component. Of course, in a polycrystalline sample with a large anisotropy both components can be absorbed by ferromagnetic resonance absorption when the sample is not completely saturated, since the random orientation of the domains which occurred in zero field has not been completely eliminated until complete saturation occurs. Fig. 10 illustrates the rotation per cm path length versus applied magnetic field for a sample of Ferramic G. Superimposed on the same figure are curves showing the absorption of the negative and positive circularly polarized components. It will be noticed that as soon as the sample is saturated, the sample becomes transparent to the negative component but almost completely absorbs the positive component.

Hence the transmitted wave at this point is almost completely circularly polarized, even though the applied magnetic field would indicate that the resonance absorption frequency was far removed from 9000 mc.

Table I gives the data taken on several ferrites at 9000 mc.

APPLICATIONS OF THE FERROMAGNETIC FARADAY EFFECT—THE MICROWAVE GYRATOR

As pointed out in the introduction, the Faraday rotation affords an anti-reciprocal phenomenon from which a microwave gyrator can be constructed. Such a gyrator is illustrated in Fig. 11 along with diagrams which help explain its action. Beneath the gyrator are construction lines which indicate the plane of polarization of a wave as it travels through the gyrator in either direction. On each diagram is a dotted sine wave which is for reference purpose only and indicates the constant plane of polarization of an unrotated wave. It is noticed that for propagation from left to right in Fig. 11, the screw rotation introduced by the twisted rectangular guide adds to the 90° rotation given to the wave by the ferrite element making a total rotation of 180°. For a wave travelling in the reverse direction, these two rotations cancel each other, producing a net zero rotation through the complete element. The unique property of the Faraday rotation becomes immediately apparent from this diagram. In the case of the rotation induced by the twisted rectangular guide, the wave rotates in one direction in going from left to right through the twisted section, and rotates in the opposite direction when it transverses the section from right to left. For the case of the rotation induced by the ferrite element, the direction of rotation is indicated by the arrow in the upper figure for either direction of propagation. The important characteristic of the element is the time phase relation between two points such as A and B in the upper diagram. It is seen with the help of the diagrams illustrating the rotating waves that the field variations are in phase at points A and B for propagation from left to right, and they are 180° out of phase for propagation from right to left. In other words the transmission line is an integral number of wavelengths long between A and B for propagation from left to right and is an odd integral number of half wavelengths long for propagation from right to left.

From the above description of the properties of the gyrator, many of its applications in microwave technology become immediately apparent. Before discussing these applications in more detail, however, it is advantageous to introduce standardized terminology and circuit symbols which apply to the gyrator and to other circuit elements derivable from it.

TABLE I

Sample Number	Material	Dimension (cm)	Applied Magnetic Field (oersteds)	Rotation/cm path	Insertion Loss (db)	Ellipticity* (db)	SWR on Input Line (db)
1	BTL $Mn_\delta Zn_{1-\delta}Fe_2O_4$	0.447 × 2.28 (length × dia.)	0	0	10.0	>50	
			245	15.6	10.3	>50	
			490	33.5	10.0	23.2	
			735	58.2	9.2	15.0	
			980	81.6	9.1	12.1	
			1225	107	9.2	10.9	
			1470	120	10	10.4	
			1715	125	11	9.3	
			1960	123	11.2	9.0	
			2206	121	11.3	7.7	
			2450	123	11.4	6.6	
			2695	—	12.4	5.0	
			2940	—	13.0	3.7	
			3185	—		3.0	
			3675	—		1.4	
2	BTL $Ni_\delta Zn_{1-\delta}Fe_2O_4$	1.36 × 2.28	0	0	0.8	>40	
			245	25	1.9	≈40	
			490	44	2.7	≈40	
			735	56	2.9	≈40	
			980	61	2.7	40	
			1225	68	2.8		
			1715	82	3.3		
			1960	85	4.9		
			2450	118	7.3	0.8	
3	Ferramic A	2.54 × 0.635	0	0	1.1	>50	0.7
			245	34.9	0.8	"	0.3
			490	43.7	0.8	"	0.3
			735	48.3	0.8	"	0.3
			980	51.1	1.0	"	0.4
			1225	54.0	1.1	"	—
			1715	57.0	1.1	"	—
			1960	60.0	1.9	"	—
			2450	63.0	3.0	35	—
			2695	64.2	3.7	—	—
4	Ferramic G	1.77 × 2.28	0	0	23.2	≫30	
			245	38	21.4	23.0	
			490	77	16.7	7.6	
			735	124	12.4	2.1	
			980	157	9.9	1.4	
			1225	170	7.7	0.7	
			1470	180	6.0	0.7	
			3430	c.p.	7.1	0.0	

* Data given is the difference in db between the major and minor components of the elliptically polarized transmitted wave.

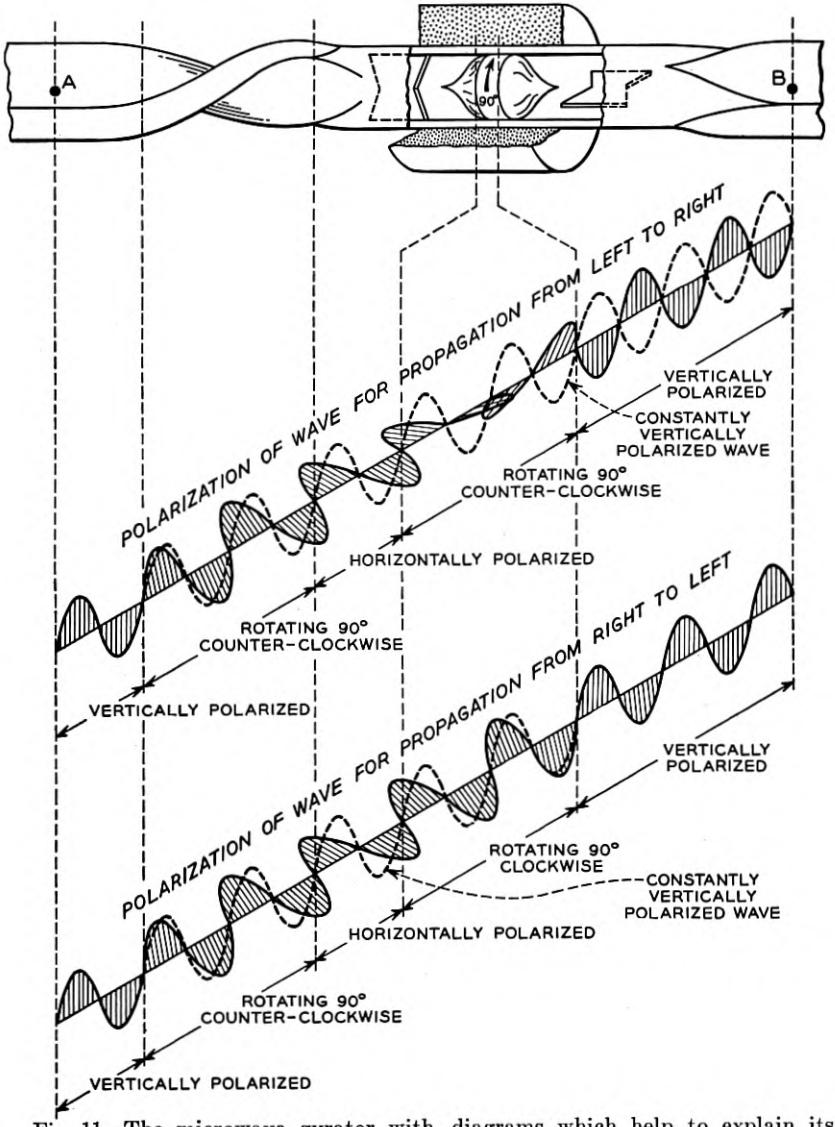


Fig. 11—The microwave gyrator with diagrams which help to explain its operation.

The "active" element of the device, the ferrite cylinder, has been termed a "Faraday Plate."

As was pointed out earlier, the fundamental property of the gyrator is the 180° phase difference introduced between the two directions of propagation through it. Thus the gyrator may be thought of as a four terminal circuit element having no phase shift for one direction of transmission, and having a 180° phase shift for the opposite direction of transmission. A convenient circuit symbol for the gyrator, which indicates this property, is shown in Fig. 12.

If the rectangular waveguides on each side of the Faraday Plate are rotated about their common axis so as to make an angle of 45° with

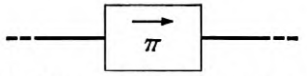


Fig. 12—Circuit symbol for gyrator.

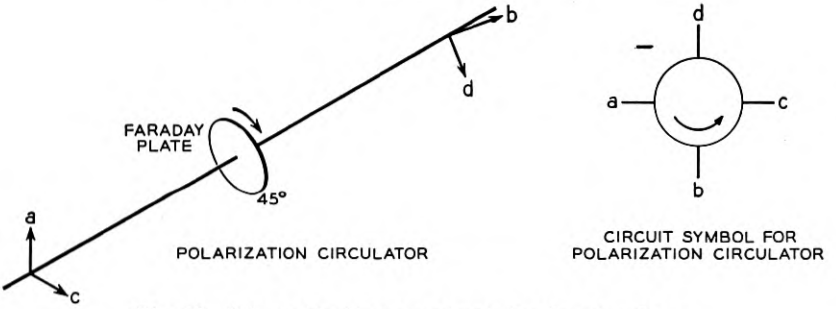


Fig. 13—Schematic diagram of polarization circulator.

each other, then a one-way transmission system can be created which is similar to Lord Rayleigh's one-way transmission system of optics, but with the important difference that this one-way transmission system does not depend upon frequency but is broad band. This one-way transmission system can be used, for example, to isolate the generator or detector from the waveguide in microwave systems. In this application it has the great advantage over the attenuators which are presently used for this purpose in that it can be made practically lossless for the direction of propagation which is desired but the reflected wave will be completely absorbed and hence more complete isolation can be effected.

A more complex and more useful circuit element, than this simple one-way transmission property would at first indicate, is obtained by adding a second connection on each side of the 45° Faraday Plate. It is suggested that this device be called a *polarization circulator*. Thus, the

polarization circulator actually has four output branches corresponding to the two different polarizations at each end. The polarizations of the four output branches are indicated in Fig. 13. It is noticed that power sent into the polarization circulator with polarization *a* is turned into polarization *b*, also *b* is turned into *c*, *c* is turned into *d*, and *d* is turned into *minus a*. This property is indicated very clearly by the circuit symbol suggested in Fig. 13, the phase inversion between arms *d* and *a* being indicated by the minus sign between the *d* and *a* arms.

Another one-way transmission system can be created by combining the gyrator with two-normal hybrids. This combination is indicated in Fig. 14. Since this device has all of the fundamental properties of the

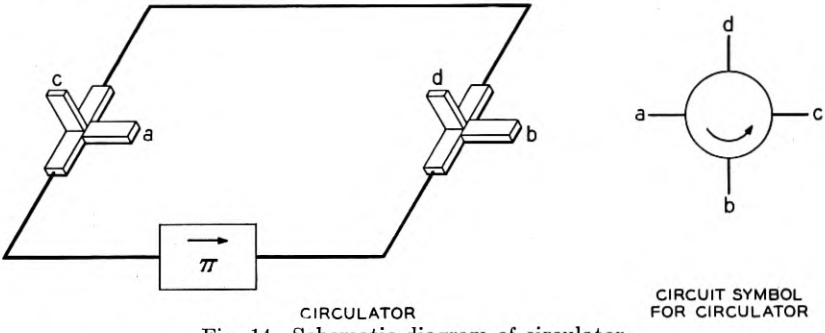


Fig. 14—Schematic diagram of circulator.

polarization circulator with the exception of the phase inversion between arms *d* and *a* it is suggested that it be called a "circulator" and the circuit symbol suggested which indicates its properties is also given in Fig. 14.

This list of applications is obviously not complete since it includes only the fundamental elements from which innumerable specific applications can be made.

In addition to the applications discussed above, which depend upon the anti-reciprocal property of the element for their operation there are several simple applications which are based only upon the fact that the amount of rotation can be controlled externally by adjusting the magnetic field. Among these uses are electrically controlled attenuators, modulators, and microwave switches.

ACKNOWLEDGMENTS

The author is indebted to a number of persons for aid in developing this circuit element. In particular, he wishes to thank A. G. Fox for

permission to use his terminology and circuit symbols, and also for the many discussions concerning the properties and uses of these microwave circuit elements. The author is also indebted to S. E. Miller for help in designing the microwave elements and to J. K. Galt for many discussions concerning the theoretical aspects of this paper. The author also wishes to extend his appreciation to J. L. Davis whose able technical assistance made possible the accumulation of much of the data presented in this paper.

APPENDIX

The equation of motion of the magnetization of a ferromagnetic material is:

$$\frac{\partial \vec{M}}{\partial t} = \gamma(\vec{M} \times \vec{H}) - \frac{\gamma\alpha}{|M|}[\vec{M} \times (\vec{M} \times \vec{H})] \tag{1}$$

where

$\vec{H}$ = internal magnetic field (oersteds)

$4\pi\vec{M}$ = magnetization of medium (gauss)

α = parameter which measures the magnitude of the damping force on the precessing dipole moment of the sample

γ = gyromagnetic ratio of the electron ($\gamma = ge/2mc$ where g is the Landé g factor for the electron).

If a ferromagnetic material is subjected to a steady magnetic field, H_a, along the z axis and if then an alternating field is applied in an arbitrary direction, Equation (1) must be solved in order to find the behavior of the magnetization of the material. To solve this problem, the following notation is introduced:

$4\pi M_z$ = magnetization of medium in absence of alternating field

H_a = externally applied steady magnetic field (oersteds)

h_x, h_y, h_z = components of applied alternating magnetic field

m_x, m_y, m_z = alternating components of magnetization

h_x^i, h_y^i, H_z^i = components of internal magnetic field

$h_x^i = h_x - N_x m_x$

$h_y^i = h_y - N_y m_y$

$H_z^i = H_a + h_z - N_z(M_z + m_z)$

N_x, N_y, N_z = demagnetizing factors of body.

Hence the magnetic field, H, occurring in Equation (1) is defined by:

$$\vec{H} = h_x^i\vec{i} + h_y^i\vec{j} + H_z^i\vec{k}$$

and

$$\vec{M} = m_x\vec{i} + m_y\vec{j} + (M_z + m_z)\vec{k}$$

In solving Equation (1), an exponential, $\exp[j\omega t]$, time dependence is assumed for the alternating magnetic field and magnetization, and if the following assumption is made:

$$h_x, h_y, h_z \ll H_a$$

it is easily shown that the alternating components of the magnetization of the medium are given by (neglecting terms of the second order in small quantities):

$$\begin{aligned} m_x &= \frac{[\gamma^2 M_z H_z^i(1+\alpha^2) + j\gamma\alpha M_z\omega]h_x^i - j\gamma M_z\omega h_y^i}{\gamma^2 H_z^{i2}(1+\alpha^2) - \omega^2 + j[2\omega\gamma\alpha H_z^i]} \\ m_y &= \frac{[\gamma^2 M_z H_z^i(1+\alpha^2) + j\gamma\alpha M_z\omega]h_y^i + j\gamma M_z\omega h_x^i}{\gamma^2 H_z^{i2}(1+\alpha^2) - \omega^2 + j[2\omega\gamma\alpha H_z^i]} \\ m_z &= 0 \end{aligned} \tag{2}$$

where:

$$j = \sqrt{-1}$$

Since

$$\vec{b} = \vec{h}^i + 4\pi\vec{m}, \tag{3}$$

it is possible by means of Equations (2) and (3) to find the relation between the alternating flux density $\vec{b}$ and the internal alternating field $\vec{h}^i$. If the ferromagnetic body is considered as being infinite, the internal fields and applied fields are equal. Hence, for this case:

$$\begin{aligned} b_x &= \mu h_x - jKh_y \\ b_y &= jKh_x + \mu h_y \\ b_z &= h_z \end{aligned} \tag{4}$$

where:

$$\mu = \mu' - j\mu''$$

$$K = K' - jK''$$

and:

$$\mu' = 1 + \frac{[\gamma^2 H_a^2(1+\alpha^2) - \omega^2][4\pi M_z \gamma^2 H_a(1+\alpha^2)] + 8\pi M_z \omega^2 \gamma^2 \alpha^2 H_a}{[\gamma^2 H_a^2(1+\alpha^2) - \omega^2]^2 + 4\omega^2\gamma^2\alpha^2 H_a^2}$$

$$K' = \frac{4\pi M_z \gamma\omega[\gamma^2 H_a^2(1+\alpha^2) - \omega^2]}{[\gamma^2 H_a^2(1+\alpha^2) - \omega^2]^2 + 4\omega^2\gamma^2\alpha^2 H_a^2}$$

$$K'' = \frac{8\pi M_z \omega^2 \gamma^2 \alpha H_a}{[\gamma^2 H_a^2(1+\alpha^2) - \omega^2]^2 + 4\omega^2\gamma^3\alpha^2 H_a^2}$$

$$\mu'' = \frac{4\pi M_z \gamma\alpha\omega[\gamma^2 H_a^2(1+\alpha^2) + \omega^2]}{[\gamma^2 H_a^2(1+\alpha^2) - \omega^2]^2 + 4\omega^2\gamma^2\alpha^2 H_a^2}$$

In order to find the behavior of a wave being propagated in this medium, it is necessary to find a solution to Maxwell's equations which are consistent with the above set of equations and in which, b, h, E, and D are of the following form:

$$\begin{aligned} \vec{b} &= \vec{b}_0 \exp[j\omega t - \Gamma(\vec{n}\cdot\vec{r})] \\ \vec{h} &= \vec{h}_0 \exp[j\omega t - \Gamma(\vec{n}\cdot\vec{r})] \\ \vec{E} &= \vec{E}_0 \exp[j\omega t - \Gamma(\vec{n}\cdot\vec{r})] \\ \vec{D} &= \vec{D}_0 \exp[j\omega t - \Gamma(\vec{n}\cdot\vec{r})] \end{aligned} \tag{5}$$

where $\vec{E}_0$, and $\vec{h}_0$ are complex vector functions of the coordinates and which satisfy the boundary conditions imposed by the waveguide. Further:

$\vec{n}$ = unit vector in the direction of propagation

Γ = propagation constant

Maxwell's equations are:

$$\begin{aligned} \nabla \times \vec{E} &= -\frac{1}{c}\frac{\partial \vec{b}}{\partial t} \\ \nabla \times \vec{h} &= \frac{1}{c}\frac{\partial \vec{D}}{\partial t} \end{aligned} \tag{6}$$

Inserting the values given in Equations (5), these become:

$$\nabla \times \vec{E}_0 - \Gamma(\vec{n} \times \vec{E}_0) = \frac{-j\omega\vec{b}_0}{c} \tag{7}$$

$$\nabla \times \vec{h}_0 - \Gamma(\vec{n} \times \vec{h}_0) = \frac{j\omega\epsilon\vec{E}_0}{c} \tag{8}$$

which can be combined to:

$$\frac{\omega^2\varepsilon}{c^2}\vec{b}_0 = \nabla \times (\nabla \times \vec{h}_0 - \Gamma\vec{n} \times \vec{h}_0) - \Gamma\vec{n} \times (\nabla \times \vec{h}_0 - \Gamma\vec{n} \times \vec{h}_0) \tag{9}$$

Writing Equation (9) in component form gives:

$$\begin{aligned}\frac{\omega^2\varepsilon}{c^2} b_x = {} & \frac{\partial^2 h_y}{\partial y \partial x} + \frac{\partial^2 h_z}{\partial z \partial x} - \frac{\partial^2 h_x}{\partial y^2} - \frac{\partial^2 h_x}{\partial z^2} - \Gamma n_x \left(\frac{\partial h_y}{\partial y} + \frac{\partial h_z}{\partial z}\right) \\ & + 2\Gamma\left(n_y \frac{\partial h_x}{\partial y} + n_z \frac{\partial h_x}{\partial z}\right) - \Gamma\left(n_y \frac{\partial h_y}{\partial x} + n_z \frac{\partial h_z}{\partial x}\right) \\ & + \Gamma^2 n_x(n_x h_x + n_y h_y + n_z h_z) - \Gamma^2 h_x\end{aligned} \tag{10}$$

$$\begin{aligned}\frac{\omega^2\varepsilon}{c^2} b_y = {} & \frac{\partial^2 h_x}{\partial y \partial x} + \frac{\partial^2 h_z}{\partial y \partial z} - \frac{\partial^2 h_y}{\partial x^2} - \frac{\partial^2 h_y}{\partial z^2} - \Gamma n_y \left(\frac{\partial h_x}{\partial x} + \frac{\partial h_z}{\partial z}\right) \\ & + 2\Gamma\left(n_x \frac{\partial h_y}{\partial x} + n_z \frac{\partial h_y}{\partial z}\right) - \Gamma\left(n_x \frac{\partial h_x}{\partial y} + n_z \frac{\partial h_z}{\partial y}\right) \\ & + \Gamma^2 n_y(n_x h_x + n_y h_y + n_z h_z) - \Gamma^2 h_y\end{aligned} \tag{11}$$

$$\begin{aligned}\frac{\omega^2\varepsilon}{c^2} b_z = {} & \frac{\partial^2 h_x}{\partial z \partial x} + \frac{\partial^2 h_y}{\partial z \partial y} - \frac{\partial^2 h_z}{\partial x^2} - \frac{\partial^2 h_z}{\partial y^2} \\ & - \Gamma n_z \left(\frac{\partial h_z}{\partial x} + \frac{\partial h_y}{\partial y}\right) + 2\Gamma\left(n_x \frac{\partial h_z}{\partial x} + n_y \frac{\partial h_z}{\partial y}\right) \\ & - \Gamma\left(n_x \frac{\partial h_x}{\partial z} + n_y \frac{\partial h_y}{\partial z}\right) + \Gamma^2 n_z(n_x h_x + n_y h_y + n_z h_z) \\ & - \Gamma^2 h_z\end{aligned} \tag{12}$$

where the subscript 0 has been dropped from all components for convenience.

If the wave in question is an infinite plane wave being propagated along the z axis, then:

$$n_x = n_y = 0, \qquad n_z = 1$$

and the components of $\vec{h}_0$ are constants, and h_z is zero. For this particular case, Equations (10), (11) and (12) become:

$$\frac{\omega^2\varepsilon}{c^2} b_x = -\Gamma^2 h_x \tag{13}$$

$$\frac{\omega^2\varepsilon}{c^2} b_y = -\Gamma^2 h_y \tag{14}$$

Equations (13) and (14) are general differential equations derivable from Maxwell's equations and do not yet contain the properties of any particular medium. In order to find the behavior of a wave travelling through an infinite ferromagnetic medium which is magnetized along the direction of propagation, it is necessary to combine these equations with Equations (4) which describe the relation between b and h in the medium. This gives:

$$(\mu h_x - jKh_y)\frac{\omega^2\varepsilon}{c^2} = -\Gamma^2 h_x \tag{15}$$

$$(\mu h_y + jKh_x)\frac{\omega^2\varepsilon}{c^2} = -\Gamma^2 h_y \tag{16}$$

The only possible solution to this set of equations is a circularly polarized wave where:

$$h_x = \pm jh_y$$

The positive sign above represents a so-called positive circularly polarized wave and the negative sign a negative circularly polarized wave. The propagation constants for these waves is given by

$$\Gamma_\pm = \frac{j\omega}{c}\sqrt{\varepsilon(\mu \pm K)} \tag{17}$$

REFERENCES

1. B. D. H. Tellegen, *Philips Research Reports*, **3**, pp. 81–101 (1948); **3**, pp. 321–337 (1948); **4**, pp. 31–37 (1949); **4**, pp. 366–369 (1949).
2. A. G. Fox, unpublished memoranda, Bell Telephone Laboratories.
3. E. M. McMillan, *J. Acous. Soc. of Am.*, **18**, pp. 344–347 (1946).
4. J. W. Miles, *J. Acous. Soc. of Am.*, **19**, pp. 910–913 (1947).
5. F. Bloch, *Phys. Rev.*, **70**, pp. 460–485 (1946).
6. H. G. Beljers and J. L. Snoek, "Gyromagnetic Phenomena Occurring Within Ferrites". *Philips Tech. Rev.*, **11**, May, 1950, pp. 313–322.
7. E. M. McMillan, *J. Acous. Soc. of Am.*, **19**, p. 922 (1947) and H. B. G. Casimir, *Nuovo Cimento*, Vol. VI, Series IX (1949).
8. W. P. Mason, unpublished memoranda, Bell Telephone Laboatories.
9. Rayleigh, *Nature*, Vol. 64.
10. H. König, *Optik*, **3**, pp. 101–119 (1948).
11. D. Polder, *Phil. Mag.*, **40**, pp. 99–115 (1949).
12. W. A. Yager, J. K. Galt, et al. *Phys. Rev.*, **80**, pp. 744–748 (1950).

Dialing Habits of Telephone Customers

BY CHARLES CLOS AND ROGER I. WILKINSON

(Manuscript received October 3, 1951)

This paper considers the behavior of customers waiting to dial calls, when dial tone is delayed. Tests were made in a panel dial central office, from which were determined: relationship between load carried by a group of line finders and the resultant dial tone delay; measures, by classes of service, of the magnitude of the generalized trunking formula's "j" factor describing the degree to which customers wait when dial tone is delayed; comparisons of observed and theoretical distributions of the number of simultaneous calls on line finder groups; and statistical accounts of the actions of customers when dial tone is delayed.

Following World War II the conversion of great quantities of manual telephone equipment to dial, and the addition of large numbers of new telephones, mostly dial, in the Bell System has directed increasing attention to those service problems peculiar to automatic operation. These problems concern chiefly the provision of adequate amounts of equipment to give satisfactory service at all times. One of the important factors affecting the amount of equipment needed is the action of the customers themselves when their calls are momentarily blocked due to these equipment shortages. The actions of subscribers whose calls are blocked due to a shortage of trunk equipment have been reported previously.[1] This paper considers the behavior of subscribers, waiting to dial calls, when dial tone is delayed.

During 1949, Bell Telephone Laboratories conducted a series of tests at the New York Telephone Company's Sterling-3 panel dial central office in Brooklyn, N. Y., with the object of increasing the knowledge available regarding subscribers' actions and their effects when dial tone is delayed.

The following principal results were obtained from the Sterling-3 tests:

1. The relationship between the load carried by a group of line finders and the resultant dial tone delay.

2. Measures, by classes of service, of the magnitude of the general-

[1] Charles Clos, "An Aspect of the Dialing Behavior of Subscribers and Its Effect on the Trunk Plant," *Bell System Tech. J.*, **27**, July 1948.

ized trunking formula's "j" factor describing the degree to which customers wait when dial tone is not immediate.[2]

3. Comparisons of observed and theoretical distributions of the numbers of simultaneous calls on the line finder groups.

4. Statistical accounts of the actions of subscribers when dial tone is delayed.

GENERAL PLAN OF THE TESTS

The general plan of the tests was developed around two specially constructed devices:

1. A 100-pen recorder capable of recording observations continuously for two hours and sensitive enough to record individual dial pulses.

2. A speed of dial tone measuring set comprising a means for manually originating test calls, and an electric timer which automatically stopped when dial tone was received.

Tests were conducted on a weekly schedule, one or two line finder groups being studied each week. Two 100-pen recorder tapes were run daily. The first was run from 10 A.M. to noon for message rate individual, flat rate individual and message rate two-party classes of service to include the morning busy periods for these customers. For coin customers the first tape was run for two hours during the noon coin busy period. The second tape was run in the afternoon from 2 to 4 P.M. for all classes of service except coin. For the coin class the second tape was run for two hours during the early evening coin busy period. Three message rate two-party tapes were run in the early evening busy period, and two coin tapes were run in the afternoons when World Series baseball games were being played in Brooklyn. A summary of the number of line finder groups observed, the number of tapes taken, the number of line finders made available and the maximum per cent dial tone delays over three seconds are given in Table I.

Except for the morning runs on the message rate individual line groups where an effort was made to maintain a fixed number of twenty line finders throughout, the number of line finders made available was selected by close observation of the flow of traffic. All studies were by half hours during which the number of line finders was held constant as far as possible. At the end of a half-hour period the number of line

[2] This formula for both finite and infinite sources was developed by R. I. Wilkinson in 1930, and appeared in the 1936 Bell Telephone Laboratories Out-of-Hour Course "The Theory of Probability as Applied to Telephone Trunking Problems." This formula for infinite sources was also developed by Conny Palm and appeared in "Etude des delais d'attente" in Erickson Technics—No. 2—1937.

finders left in service was adjusted in order to obtain a reasonable number of dial tone delays in the next half hour without producing a severe reaction from the subscribers served by the line groups under study.

The data recorded on the tapes showed continuously the busy or idle conditions of certain circuits associated with the line groups under study. In some cases the receipt of dial pulses and the operation of registers were also recorded. These circuits included line finders, a few subscriber lines, trip circuit sub-groups, the all trunks busy register, the peg count register and the speed of dial tone measuring device. In

TABLE I

	Number of Line Finder Groups Observed	Number of Tapes	Number of Line Finders Made Available	Maximum Percent Dial Tone Delays Over 3 Seconds
Message rate individual	11			
Morning tapes		25	19 to 20	53.3
Afternoon tapes		29	10 to 15	55.6
Message rate two-party	3			
Morning tapes		8	18 to 19	40.0
Afternoon tapes		12	10 to 12	88.6
Evening tapes		3	19 to 21	71.2
Flat rate individual	2			
Morning tapes		9	14 to 33	48.9
Afternoon tapes		10	10 to 17	35.6
Coin	3			
Morning tapes		18	30 to 39	71.1
Afternoon tapes		2	30 to 39	35.6
Evening tapes		10	25 to 39	68.9

addition the busy and idle conditions of a sample of senders was observed in order to note the general load level on the senders.

RELATIONSHIP BETWEEN LOAD CARRIED AND PER CENT DIAL TONE DELAY

One of the principal objectives of these tests was to establish as far as possible the relationship between the average load carried by a line-finder group and the corresponding dial tone service when there is a shortage of line finders but not of senders.[3] The average load carried was obtained by making a switch count every thirty seconds of the number of line finders busy as indicated by the 100-pen recorder tape. The dial tone tests were made with the speed of dial tone measuring device. Forty-five dial tone tests were made each half hour for each

[3] The restriction of avoiding dial-tone delays due to a sender shortage was to eliminate a factor external to the line-finder group and to make the results of the tests applicable to both common-control and non-common-control type dial systems.

line group studied. Additional dial tone tests were made on all other line groups in the office as a check that the delays experienced on the line groups under study were not due to a sender shortage. The sender data on the tapes were also used for this purpose. Figs. 1(a) to 1(d), and 2(a) to 2(d), inclusive, show for various amounts of load carried, the per cent of dial tone tests encountering delays greater than three and greater than ten seconds for half-hour study periods for the most frequent number of line finders in the tests for each class of service.

Plotted on each of these figures is a theoretical fitting dial tone tester delay curve computed for the indicated dial tone delays and for the following j factor values in the generalized trunking formula, determined in a manner to be explained later:

Class of Service	j factor
MRI	6.6
MR 2-party	5.8
FRI	6.5
Coin	2.1

To indicate the effect of varying j, several curves have been added to Fig. 1(a). Selections of curves for $j = 0$, 1 and ∞ (which correspond to the three commonly used infinite source congestion formulae, Erlang C, Poisson and Erlang B when adapted to the tester's delay problem) are shown. It is clear that with the wide differences in delays which they give for specified loads carried, it is highly desirable to select that j formula for engineering use which most nearly describes the customer actions in any situation being dealt with. In the field of curves shown on Fig. 1(a), the one labelled $j = 6.6$ was derived in a logical manner from the data, and shows an agreeably satisfactory fit. For example, during a heavy load period when, say, 20 per cent of a dial tone tester's calls are meeting delays greater than 3 seconds, an actual average load of 16.6 erlangs (as shown by the $j = 6.6$ curve) would likely be carried. (Load in erlangs equals average number of simultaneous calls.) The Erlang C ($j = 0$) and Poisson ($j = 1$) theories would indicate the presence of loads of 15.6 and 16.0 erlangs, respectively, figures clearly too small for the circumstances shown by the data of Fig. 1(a). On the other hand, use of the Erlang B ($j = \infty$) theory would predict a considerably larger load carried, about 17.9 erlangs, than one would probably be justified in assuming here for engineering purposes.

By grouping the dial tone delay data by bands of load carried, relationships of per cent of test calls encountering varying dial tone delays

up to twelve seconds were obtained. These data are shown on Figs. 3 to 6. Fig. 3 is for the message rate individual class of service with twenty line finders available. The data on this figure correspond to those on Figs. 1(a) and 1(b). Fig. 4 is for the message rate two-party class of service with ten line finders and corresponds to Figs. 1(c) and 1(d). Fig. 5 is for the flat rate class of service with ten line finders and cor-

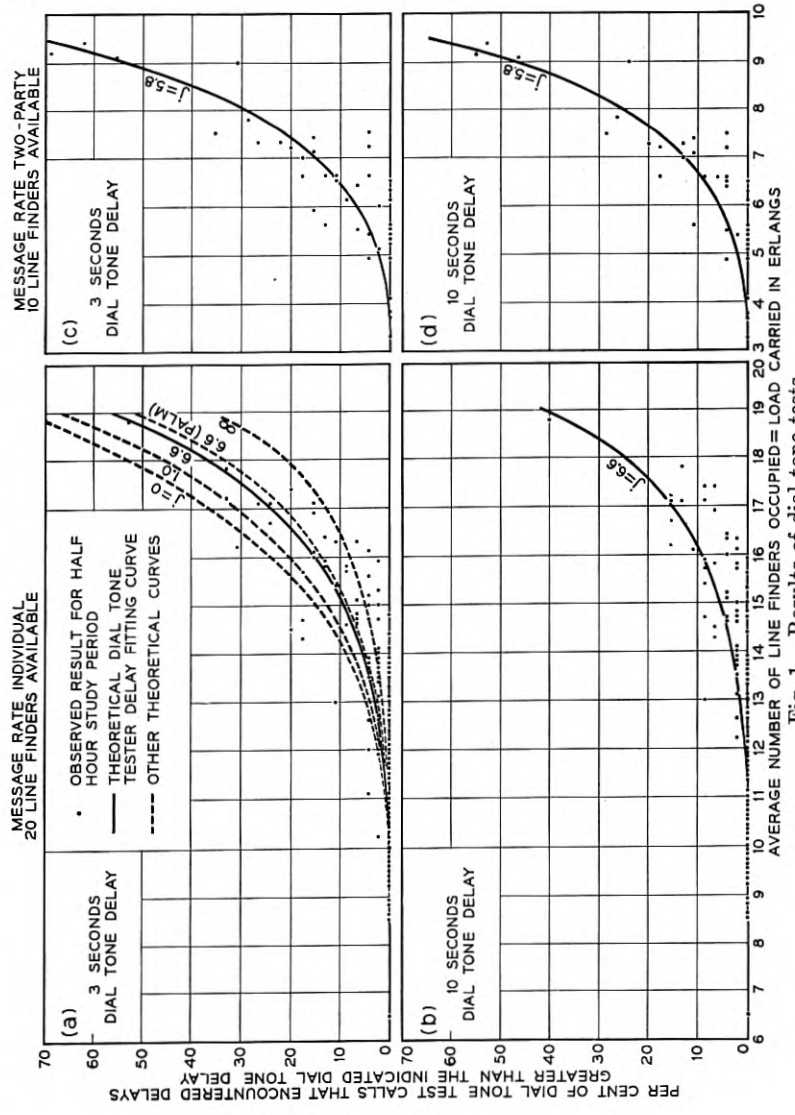


Fig. 1.—Results of dial tone tests.

responds to Figs. 2(a) and 2(b). Fig. 6 is for the coin class of service with 34 line finders and corresponds to Figs. 2(c) and 2(d).

Plotted on Figs. 3 to 6 are theoretical fitting dial tone tester delay curves, curves A, determined by means of the following formulae:

1. The generalized trunking formula for determining the proportion of calls that encounter congestion, i.e., find all line finders busy.

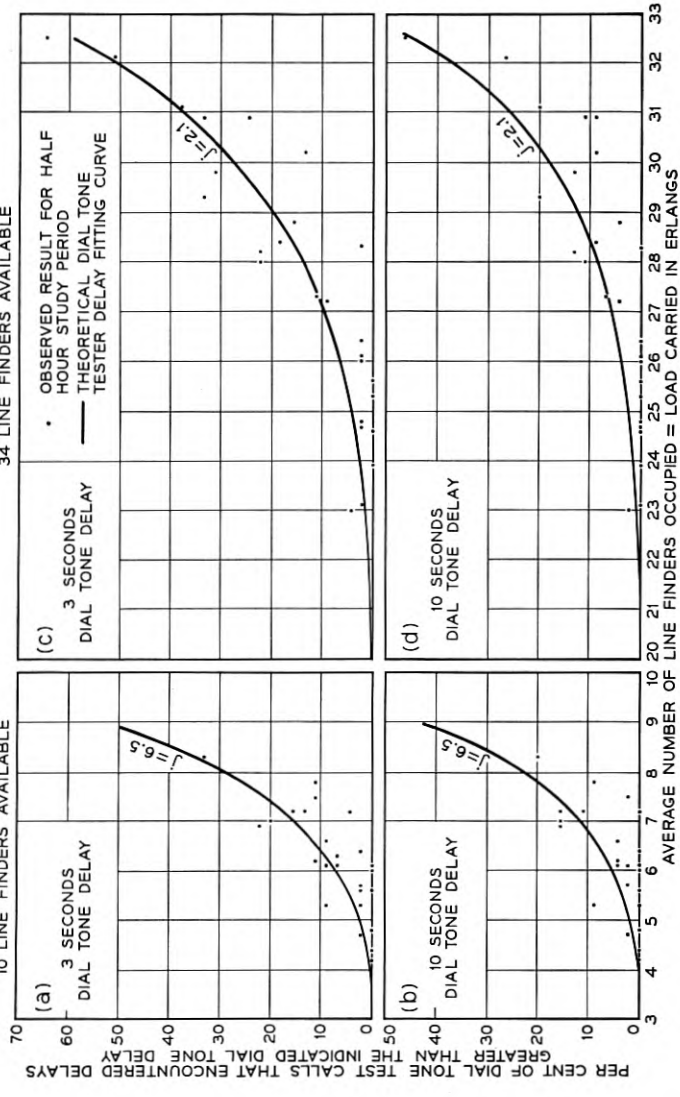


Fig. 2—Results of dial tone tests.

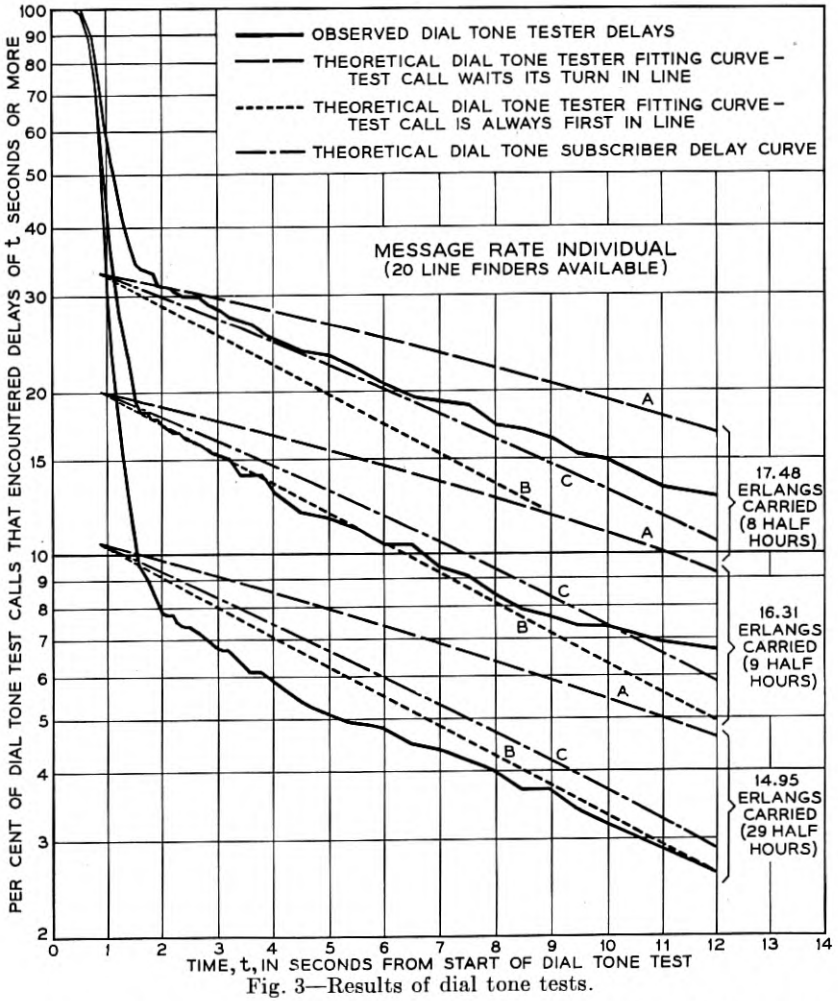


Fig. 3—Results of dial tone tests.

2. A delay formula[4] for determining the proportion of those test calls encountering congestion which will have a delay in obtaining dial tone of at least time t.

Some of the theoretical aspects of these formulae are considered in the two sections that follow.

GENERALIZED TRUNKING FORMULA

The generalized trunking formula combines in one expression the various assumptions underlying the Erlang B, Poisson, and Erlang C

[4] A development by John Riordan paralleling that of Conny Palm's.

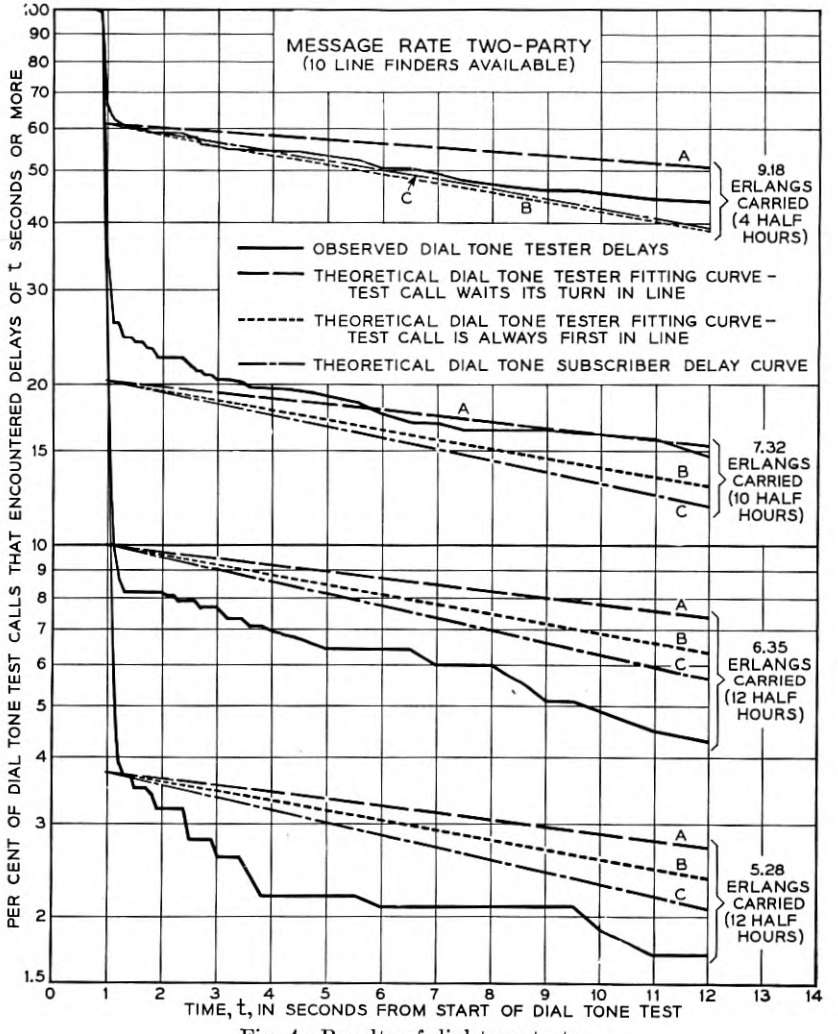


Fig. 4—Results of dial tone tests.

trunking formulae and a large field of intermediate assumptions regarding the disposition of calls which fail to obtain immediate service. These assumptions are given in Table II on page 42.

One method for developing the generalized trunking formula[5] is to consider the probabilities of the existence of certain states and to determine the number of transitions during some convenient interval of

[5] In this article only the unlimited sources trunking formula is considered.

time from one state to another. By equating certain of these transitions, a series of simultaneous equations evolves, which when solved yields one overall expression. Of interest is the development of the transition equations. Thus for a case of c line finders arranged in a simple group, let $f(x)$ represent the probability that x (where $x < c$) line finders are occupied and $f(x + 1)$ represent the probability that $x + 1$ line finders are occupied. Let n be the average number of calls offered to the line finders during a long interval of time, T. Let T be the unit of time and h be the average holding time per call measured in terms of T. Over a

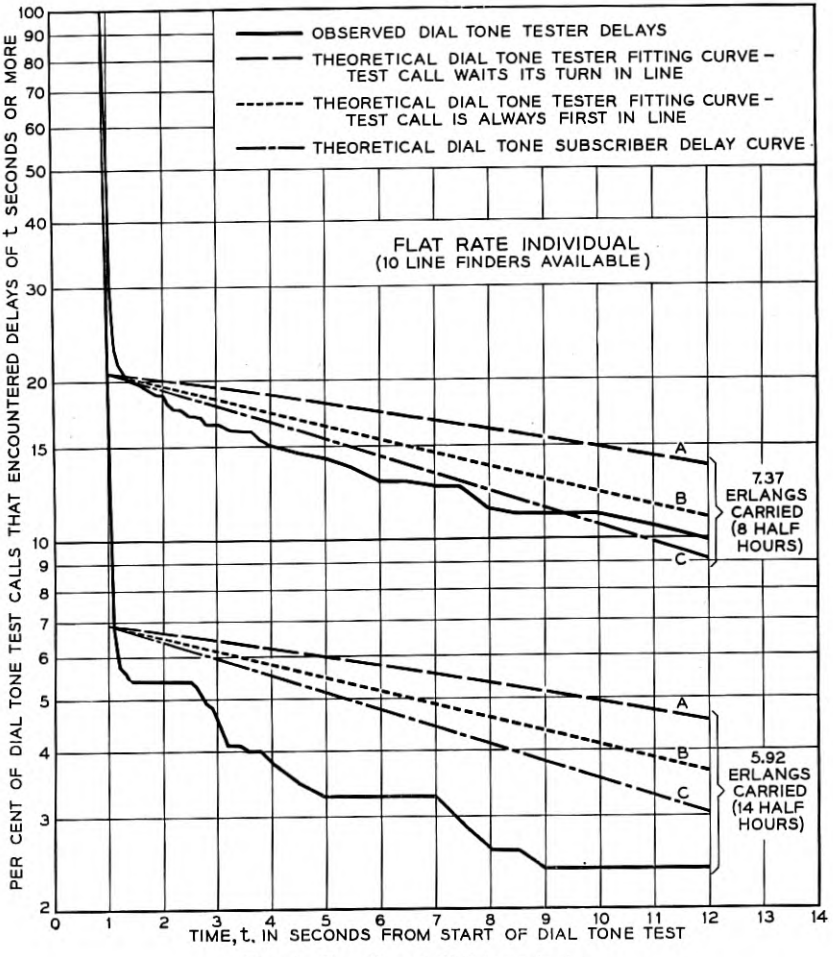


Fig. 5—Results of dial tone tests.

period T the number of transits from state x to state $x + 1$ must equal (or differ by no more than one) the number of transits in the reverse direction. That is:

$$nf(x) = \frac{x + 1}{h} f(x + 1) \tag{1}$$

It may be noted that $nh = a$, where a is the average offered traffic load in erlangs.

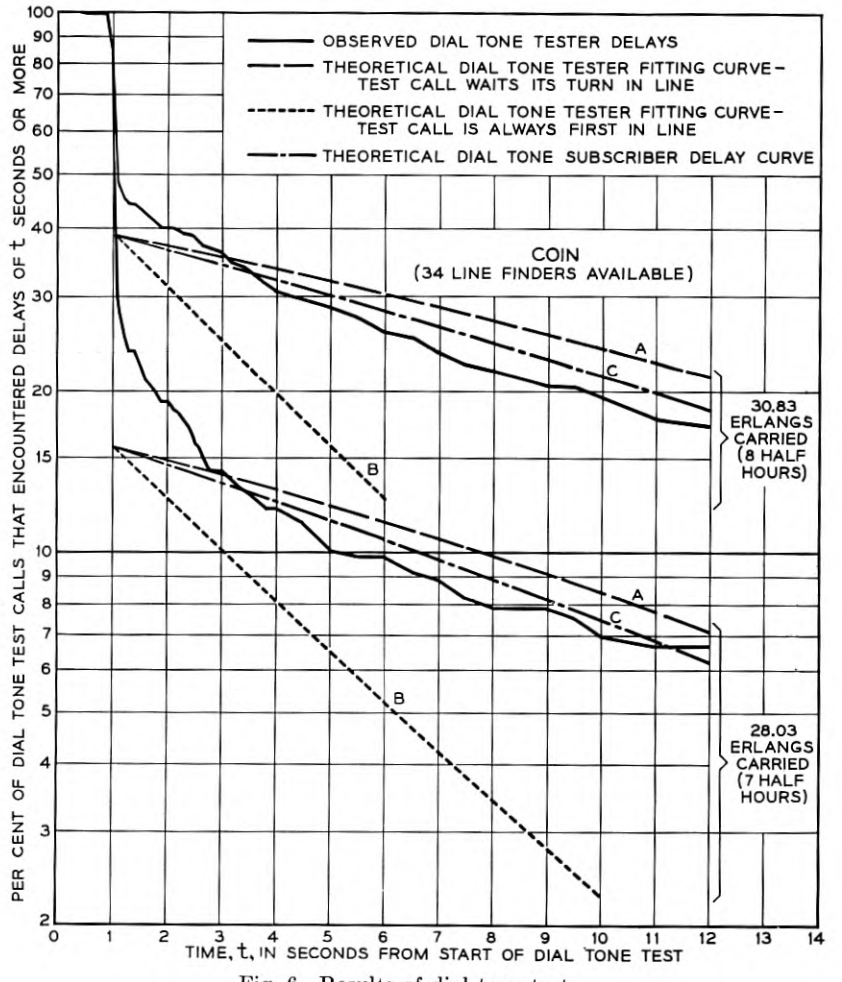


Fig. 6—Results of dial tone tests.

TABLE II

Formula	Assumption Concerning the Disposal of Calls that do not Obtain a Line Finder Immediately
Generalized.....	Waiting calls are cleared out at a rate j times the rate at which calls are terminated when served by the line finders.
Erlang B.....	Calls are cleared out of the system immediately, that is no calls wait (j = ∞).
Poisson.....	Waiting calls are cleared out at a rate equal to that with which calls are terminated when served by the line finders (j = 1).
Erlang C.....	Calls wait until served (j = 0).

When $x \geqq c$ a new situation is encountered. "c" calls are engaged in conversation and $x - c$ calls are waiting for service. If the waiting calls are forced to wait for an unduly long period of time so that in effect they are being denied service, it can be expected that they will wait for some average period, say H, and then abandon their attempts. On this basis the corresponding equation is:

$$nf(x) = \frac{c}{h} f(x+1) + \frac{(x+1-c)}{H} f(x+1) \tag{2}$$

It has been assumed in the above equations that the distribution of the holding times is exponential, an assumption which is found in most local systems to be reasonably justified. The distribution of the waiting times is also taken to be exponential. By introducing a factor j, where $j = h/H$, equation (2) can be written in the simpler form:

$$af(x) = [c + j(x + 1 - c)]f(x+1) \tag{3}$$

Solving this system of simultaneous equations, we obtain:
when $x < c$,

$$f(x) = \frac{a^x}{x!} f(0) \tag{4}$$

when $x \geqq c$,

$$f(x) = \frac{a^x}{c!(c+j)(c+2j)\cdots\cdots\cdots[c+(x-c)j]} f(0) \tag{5}$$

where

$$f(0) = \left(\sum_{x=0}^{x=c} \frac{a^x}{x!} + \sum_{x=c+1}^{x=\infty} \frac{a^x}{c!(c+j)(c+2j)\cdots[c+(x-c)j]} \right)^{-1} \tag{6}$$

The probability of a call encountering congestion, which is equivalent to the probability of a call having a delay greater than zero units of time is:

$$P(>0) = \sum_{x=c}^{x=\infty} f(x) \tag{7}$$

DELAY FORMULA FOR THE DIAL TONE TESTER

The probability that a dial tone test call encounters congestion is given by expression (7). Once a test call has encountered congestion it will experience a delay depending upon a number of variables. The assumptions underlying the dial tone tester formula are:

1. A dial tone test call when encountering a delay waits until served.
2. A dial tone test call does not add to the load offered and carried by the line finders.
3. Upon encountering a delay, a dial tone test call is served in the order of its arrival with respect to all other waiting calls. For example, if the test call finds three other calls waiting, it waits fourth in line.

Under the third assumption as calls drop out, due to conversations terminating on the occupied line finders or due to waiting calls abandoning their requests for service, the test call advances from an initial position of say fourth in line to third in line, then to second, then to first in line, and finally is served. The overall delay distribution of the test calls depends therefore upon the number of calls they find waiting ahead of them. The delay distribution for each such number must be weighted by the probability of its occurrence in order to obtain the overall distribution. The delay distribution for a test call which finds zero calls waiting is:

$$p_0(>t) = \exp(-ct/jH) \tag{8}$$

The probability is $f(c)$ that a call made at random will find all line finders busy with no calls waiting. Hence the weighted delay distribution $P_0(>t)$, is:

$$P_0(>t) = f(c)p_0(>t) = f(c)\exp(-ct/jH) \tag{9}$$

The delay distribution for a test call which finds one call waiting ahead of it is:

$$p_1(>t) = [1 + c/j - (c/j)\exp(-t/H)]\exp(-ct/jH) \tag{10}$$

The probability is $f(c+1)$ that a call made at random will find all

line finders busy and one call waiting. Hence the weighted delay distribution, $P_1(>t)$, is:

$$P_1(>t) = f(c+1)[1 + c/j - (c/j)\exp(-t/H)]\exp(-ct/jH) \quad (11)$$

In the general case $p_n(>t)$ is given by the following formula:

$$p_n(>t) = F_{n+1}(t)\exp(t/H) \quad (12)$$

where $F_{n+1}(t)$ is given by Conny Palm.[6] The over-all delay distribution is then:

$$P(>t) = P_0(>t) + P_1(>t) + P_2(>t) + \cdots \quad (13)$$

By making appropriate substitutions and summing the result, expression (13) becomes:

$$P(>t) = \frac{a^c}{c!} f(0)\left[1 + \frac{a\exp(-t/H)}{c+j} + \frac{a^2\exp(-2t/H)}{(c+j)(c+2j)} + \cdots\right] \exp\left[-ct/jH + (a/j)[1 - \exp(-t/H)]\right] \quad (14)$$

Expression (14) is equivalent to that of Riordan involving two incomplete gamma functions as follows:

$$P(>t) = P(>0)\frac{\gamma[c/j, (a/j)\exp(-t/H)]}{\gamma(c/j, a/j)} \quad (15)$$

where the incomplete gamma function,

$$\gamma(N, x) = \int_0^x x^{N-1}e^{-x}\,dx \quad (16)$$

The theoretical dial tone tester delay curves shown on Figs. 1(a) to 1(d), 2(a) to 2(d), and 3 to 6 were computed from expression (14), using the following values of j and H for the classes of service studied, these values being determined in a manner explained later:

Class of Service	j factor	H
MRI	6.6	24 seconds
MR 2-party	5.8	42 seconds
FRI	6.5	27 seconds
Coin	2.1	74 seconds

On Figs. 1(a), 1(c), 2(a), and 2(c), which show the per cent of dial tone tests encountering delays greater than three seconds for various amounts

[6] Equation 53, loc. cit.

of load carried, it may be noted that most of the theoretical dial tone tester delay curves are in close agreement with the observed data, with a tendency perhaps to be slightly high. On Figs. 1(b), 1(d), 2(b), and 2(d), which are for dial tone delays greater than ten seconds, it may be noted that the theoretical curves have a slightly stronger tendency to lie on the high side of the observed data. On Figs. 3 to 6 the theoretical dial tone tester delay, curves A, again lie in the proximity of the curves of the observed data, with a tendency to lie higher than these latter curves, especially at the ends where the dial tone delays are greatest. Among the factors which account for this discrepancy are:

1. A feature is present in panel line finder circuits for momentarily releasing trip circuits with waiting calls to prevent the orphaning of calls under certain trouble conditions. The release occurs after a call has been waiting from 5 to 12 seconds and reoccurs every 7 seconds thereafter. When such a release occurs the call yields whatever waiting preference it may have had to a subsequently placed call which is not yet affected by such a release. The dial tone test calls did not wait beyond 12 seconds. Hence for these test calls there was only one possibility of such a release and for many of them the release occurred near the end of their waiting period. Hence they were more likely to gain preference over other calls than to lose their preference.

2. Subscribers while waiting for dial tone frequently become impatient and proceed to flash (move their switchhook up and down). While flashing, a subscriber may lose preference to a subsequently placed test call (the latter of course does not flash).

3. Many subscribers fail to observe dial tone and proceed to dial. During such dialing, a subscriber may lose preference to a subsequently placed test call.

4. Line finders serve a large proportion of call attempts of short holding time whose presence may militate against the occurrence of the longer delays. In connection with the measurement of the j factor, the following proportions of call attempts and average holding times were noted on which no dialing occurred or where no more than two digits were dialed.

Class of Service	Proportion of Attempts with No Peg Counts	Average Holding Time
MRI	35.2%*	4.4 seconds*
MR 2-Pty.	33.3%*	5.8 seconds*
FRI	25.1%	5.8 seconds
Coin	9.8%	8.3 seconds

* Partly estimated

The individual contributions of these four factors to discrepancies between theory and observation are not easy to assess. The first three explain a tendency for test calls to get ahead of calls already waiting for dial tone. On Figs. 3 to 6, inclusive, additional theoretical dial tone delay curves, curves B, for the case where a dial tone tester always gets first in line are shown. Even these curves tend to lie above the curves of the observed data on Figs. 4 and 5 where ten line finders were available; they more nearly agree with the observed data on Fig. 3 where twenty line finders were available, and they lie below the observed data on Fig. 6 where 34 line finders were available. This is an indication of the fact that with higher traffic loads (which occurred on the larger line finder groups) a test call will encounter more competition from other calls and therefore will have a lesser chance of gaining precedence over all of the other calls. The fourth factor indicates that the call attempts served on line finders consist of two distinct holding time universes and not just one, as was assumed in the development of the dial tone tester formula. The effect of the presence of both a short and long holding time universe of calls would be to introduce a change of slope in the delay curves which may be seen in Figs. 3 to 6 to be at about $t = 4$ seconds. There is reason to believe that the same cause may have been responsible for the tendency of the observed delay curves to fall away from the theoretical at the lower levels of load carried.

Due to the reasons given above and to the fact that the dial tone delay observations were made by the test call method, the above results may not directly describe service from the customer's point of view. Conny Palm has developed the following formula which gives a slightly different measure of customers' dial tone service. It indicates the proportion of calls which have neither received dial tone nor have dropped out at time t.

$$P(>t) = P(>0)\,\frac{\gamma[c/j,\,(a/j)\,\exp\,(-t/H)]}{\gamma(c/j,\,a/j)}\,\exp\,(-t/H) \qquad (17)$$

Curves for this formula are shown plotted on Fig. 1(a) and at C on Figs. 3 to 6. They are quite close in many cases to the observed dial tone tester results. It would appear that a sufficiently good estimate of the customer's dial tone service, whatever its precise definition, can be obtained by the dial tone tester method.

Recently revised tables for the capacity of step-by-step line finders have been published for Bell System use based on Palm's formula using a factor of $j = 5$. This was selected as being slightly conservative for

most applications after reviewing the above Sterling-3 results and other line finder data collected in step-by-step offices.

MEASUREMENT OF THE j FACTOR BY CLASSES OF SERVICE

As indicated previously, the data recorded on the tapes showed the states of being busy or idle and of changes in these states for line finders and the associated trip circuits. A fully equipped line finder group of 400 lines has ten trip circuits each of which serves two sub-groups of twenty subscriber lines in the following manner. When a line originates a call its line relay is operated. This causes a ground to appear on a lead which is common to all twenty line relays in the sub-group and starts a line finder hunting for the calling subscriber's line. As soon as this hunt is completed the cutoff relay associated with the calling line operates and disconnects the line relay, removing the ground (unless, of course, another line in the sub-group has originated a call in the meantime). During periods of overload when line finders are not immediately available, the ground due to a single subscriber will persist until:

1. A line finder is obtained, or
2. The subscriber abandons the attempt, or
3. The subscriber receives an incoming call which operates the cutoff relay.

The twenty leads from the trip circuit sub-groups were brought out to the pen recorder and a record taken of the grounds that occurred on each lead. Except for the possibility that more than one subscriber is waiting for service at the same time on a given trip circuit sub-group, the record of the occurrences of the grounds gives a substantially accurate[7] record of the demands for service and of the number of calls waiting for service. Hence an analysis of the events occurring on the trip circuit sub-groups and on the line finders as recorded on the tapes gives a means for determining H. The quantity H was introduced in equation (2) in the term

$$\frac{(x + 1 - c)}{H} f(x + 1) \tag{18}$$

For convenience in the ensuing discussion this term will be replaced by

[7] To obtain absolute accuracy would require the use of a pen recorder with one pen for each of the 400 subscribers served on a line finder group plus one for each line finder.

the equivalent expression:

$$N_y = \frac{y}{H} f(z + y) \tag{19}$$

where N_y = The average number of waiting calls that drop out per unit of time during the state $(z + y)$.
y = The number of waiting calls.
z = The number of line finders occupied with calls.
$1/H$ = A measure of the rate at which calls tend to abandon waiting.
$f(z + y)$ = The proportion of time that the state $(z + y)$ exists.

On the tapes we can measure $f(z + y)$ and count N_y. Hence H can be determined. The result is a statistical quantity subject to many chance factors. In the actual analysis of a tape, the composite average value of H was determined for all possible observed states where calls were waiting. By an analogous process, the composite average value of h for all possible observed states where calls were being served by line finders was determined. Also as a side computation, a composite average value of h' for calls that were served by line finders but for which no peg counts were scored was determined. This value of h' is included in h on the basis that data for engineering line finders consist of estimated calls based on peg counts and of holding times which include an allowance for these short holding time calls. The average values of H, h and of the j factor for the four classes of service studied are given in Table III.

The results for H by individual half hours and by various percentages of dial tone delays greater than three seconds are shown on Figs. 7(a) to 7(d) respectively for the four classes of service. On some of these figures an upward bulge may be noted in the center. This is not considered to be characteristic of the habits of the subscribers but is the overall effect resulting from a number of arbitrary rules followed in making the analysis in order to simplify the work and to offset par-

TABLE III

	Average Values in Seconds		$j = h/H$
	H	h	
Message rate individual	24	159	6.6
Message rate two-party	42	243	5.8
Flat rate individual	27	176	6.5
Coin	74	153	2.1

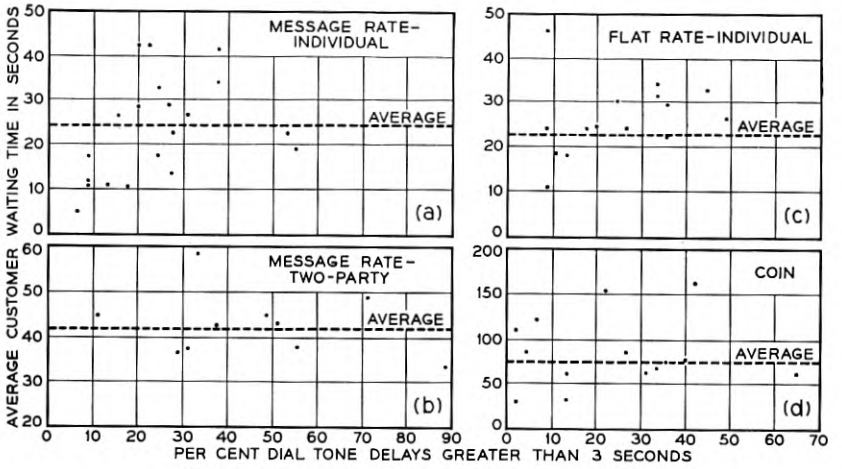


Fig. 7—Average customer waiting time (H).

tially the effect of occasionally having two or more calls waiting on one trip circuit sub-group. The rules and the reasons for them will be described with the aid of Fig. 8, which shows a hypothetical section of one of the tapes. The rules were as follows:

1. Initial Overlap

Referring to Fig. 8, at t_1 a subscriber has initiated a request for service. At t_2 a line finder rises to serve the subscriber. At t_3 the subscriber receives service. This case is typical of a subscriber receiving prompt dial tone service.

The span from t_1 to t_2 was difficult to measure accurately because, for the usual case, it was about the same as the maximum error due to misalignment of the recorder pens. It was not measured unless the combined span from t_1 to t_3 exceeded one second.

The span from t_2 to t_3 involves an overlap, it represents a period when a line finder is busy hunting for the terminal of the subscriber who originated the request for service. It also represents a period when a subscriber is waiting for service. In the analysis this span was treated as a case where a line finder was busy with a call and not as a call waiting for service.

If the span from t_2 to t_3 and all similar cases had been treated as calls waiting for service and if in addition all spans from t_1 to t_2 which were not measured had also been treated as calls waiting for service, the average values for H would have increased slightly for each class of service.

2. Three Second Rule for the Bridging of Calls

Referring to Fig. 8, again, at t_5 a request for service is originated on trip circuit 5 and at t_7 this request is withdrawn. At t_{10} apparently a new request for service is initiated which is then withdrawn at t_{11}. From manual service observations it is found that subscribers often flash when dial tone is slow. A few pens were used to observe individual subscribers, and Fig. 9 shows a case where a subscriber made several flashes when his tone was slow. When a subscriber flashes it appears as though he

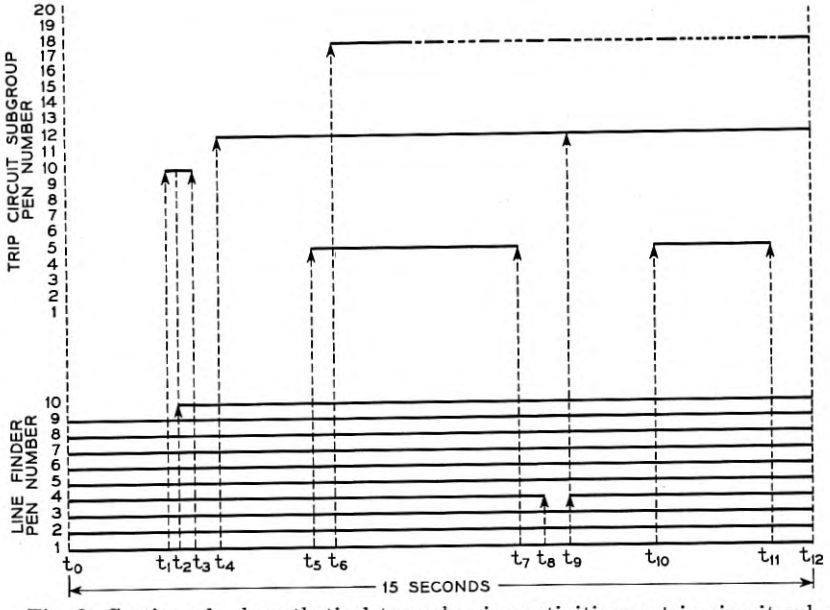


Fig. 8—Section of a hypothetical tape showing activities on trip circuit sub-groups and on line finders.

were making several bids for service. Actually he is making only one real bid. On the trip circuit sub-group pens it was generally impossible to distinguish between flashes and requests for service by two or more subscribers. To resolve this problem many observations of subscriber lines recorded on the tapes were examined from which it was concluded that no great error would result if a break in the demand for service on a trip circuit sub-group of less than three seconds were considered as a flash and was to be bridged, and a break greater than three seconds was to be considered as the termination of one call attempt and the start of another

3. Treatment of Cases Where Two or More Calls Were Found to be Waiting on One Trip Circuit Sub-Group

The occurrence of several calls waiting on one trip circuit was occasionally noted in the analysis. Referring to Fig. 8, a case is shown on trip circuit sub-group 12. At t_9 a line finder is seized. Trip circuit sub-group 18 shows that a subscriber is dialing before tone. The appearance of dial pulses on this trip circuit indicates that only one subscriber is demanding service otherwise the dialing would not show. Trip circuit sub-group 12 however appears to have two or more requests for service. One of these requests for service began at t_4. The start of the second request occurred somewhere between t_4 and t_9, perhaps half-way between. At t_9, one of the requests was served by a line finder. To simplify the handling of such cases, the assumption was made that the first attempt started at t_4 and ended at t_9 and the second request started

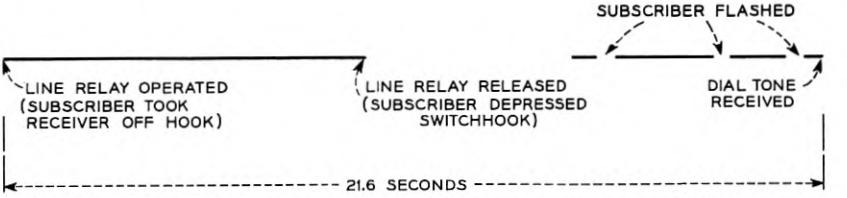


Fig. 9—Example of a customer flashing for dial tone (Tape made October 18).

at t_9. The effect of this is to understate by an indeterminate amount the average value of H for each class of service. This understatement should be noticeable for the higher degrees of overload because the occurrence of several calls on one trip circuit sub-group is then most likely to occur.

The effect of several calls simultaneously waiting in a trip circuit sub-group and of one or more calls dropping out is to overstate the magnitude of H. For instance if two simultaneous call attempts of five seconds each overlap for one second and both attempts are abandoned, the apparent average waiting period is nine seconds, whereas it should be five seconds. It is believed that the above three rules tend to create understatements which roughly balance this type of overstatement.

DISTRIBUTION CURVES OF SIMULTANEOUS CALLS

The detailed analysis of the tapes provided distributions of simultaneous calls. For each class of service studied these distributions can be compared with theoretical distributions derived from the generalized trunking formula using the j factors developed in the analysis. Several

such comparisons are shown on Figs. 10 and 11. The agreement is quite good in most cases.

SUBSCRIBER DIALING HABITS AS OBSERVED WITH A MONITORING CIRCUIT ON A SENDER WITH INDUCED DIAL TONE DELAYS[8]

As a separate study a series of tests was made by means of a monitoring circuit on one of the senders serving in common the subscribers in the Sterling-3 and Main 2 central offices, for the purpose of obtaining further information on subscriber dialing characteristics under overload conditions. A large amount of data was collected on the time intervals from the seizure of the sender to the first action taken by subscribers when encountering dial tone delays, the latter being introduced under the control of the observer.

The monitoring circuit was wired to a particular sender in a group of 100 serving all classes of subscribers. When the circuit was in use, the only irregularity introduced was that the dial tone could be delayed even though the sender was actually available to the subscriber. The delay did not affect the sender in its functions if the subscriber elected to dial before tone.

The sender monitoring circuit provided the following four features:

1. A receiver was bridged across the tip and ring leads in the sender so that an observer could hear certain actions taken by a subscriber connected to the sender. The sender was of course disconnected before conversation.

2. The observer was able to preselect one of several intervals by which dial tone was delayed on successive calls served by the sender. This was accomplished with a capacitance-resistance-vacuum tube circuit.

3. By means of a timer which started when the sender was seized, the observer was enabled to note elapsed time intervals to the occurrence of the various actions of the subscribers. The reading of the time of the first action of a subscriber had to be made when the second hand was in motion, which introduced certain errors later to be discussed.

4. By means of colored lamps the observer was able to classify all calls observed as being message rate, flat rate or coin.

During the sender dial tone delay tests, observations were made only during the afternoons when the flow of traffic was light and the probability of a subscriber obtaining a delay before reaching the sender was a minimum.

[8] Based on an unpublished report by W. A. Reenstra.

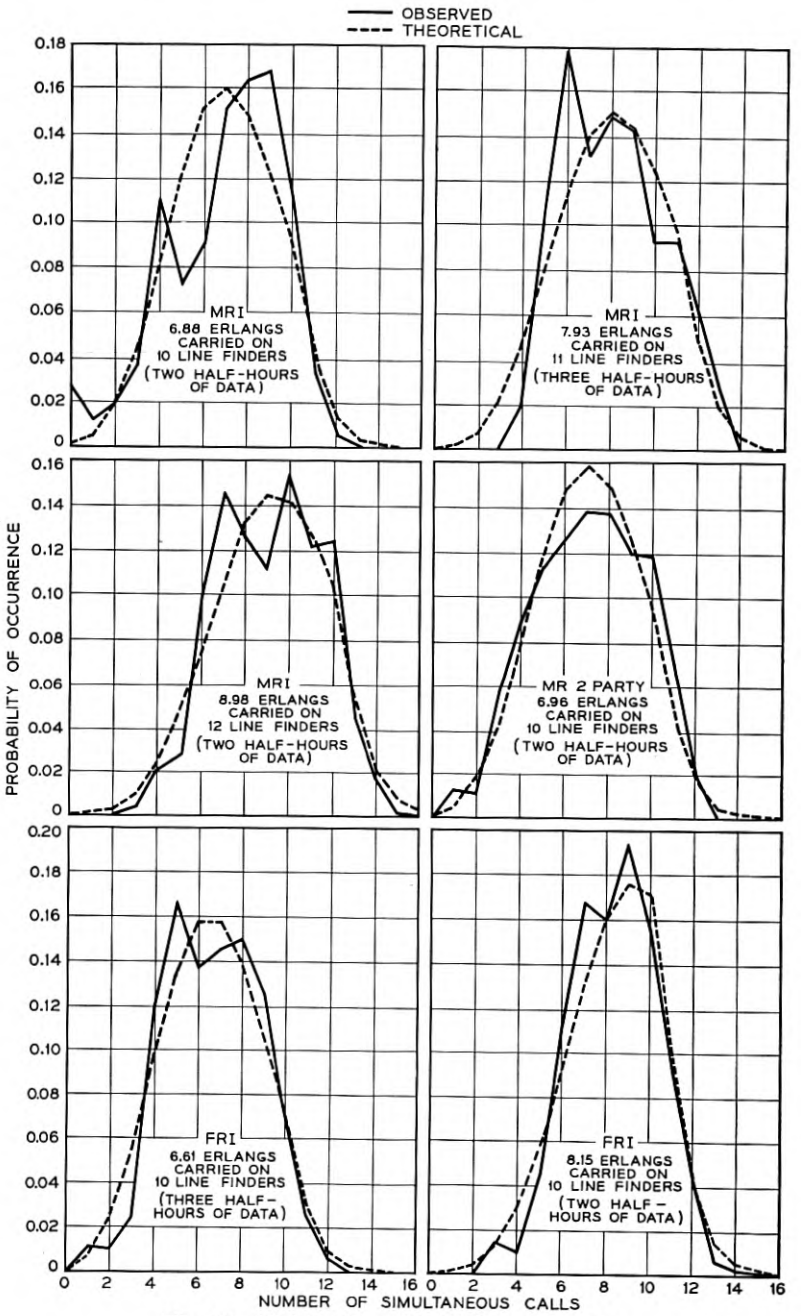


Fig. 10—Distributions of simultaneous calls.

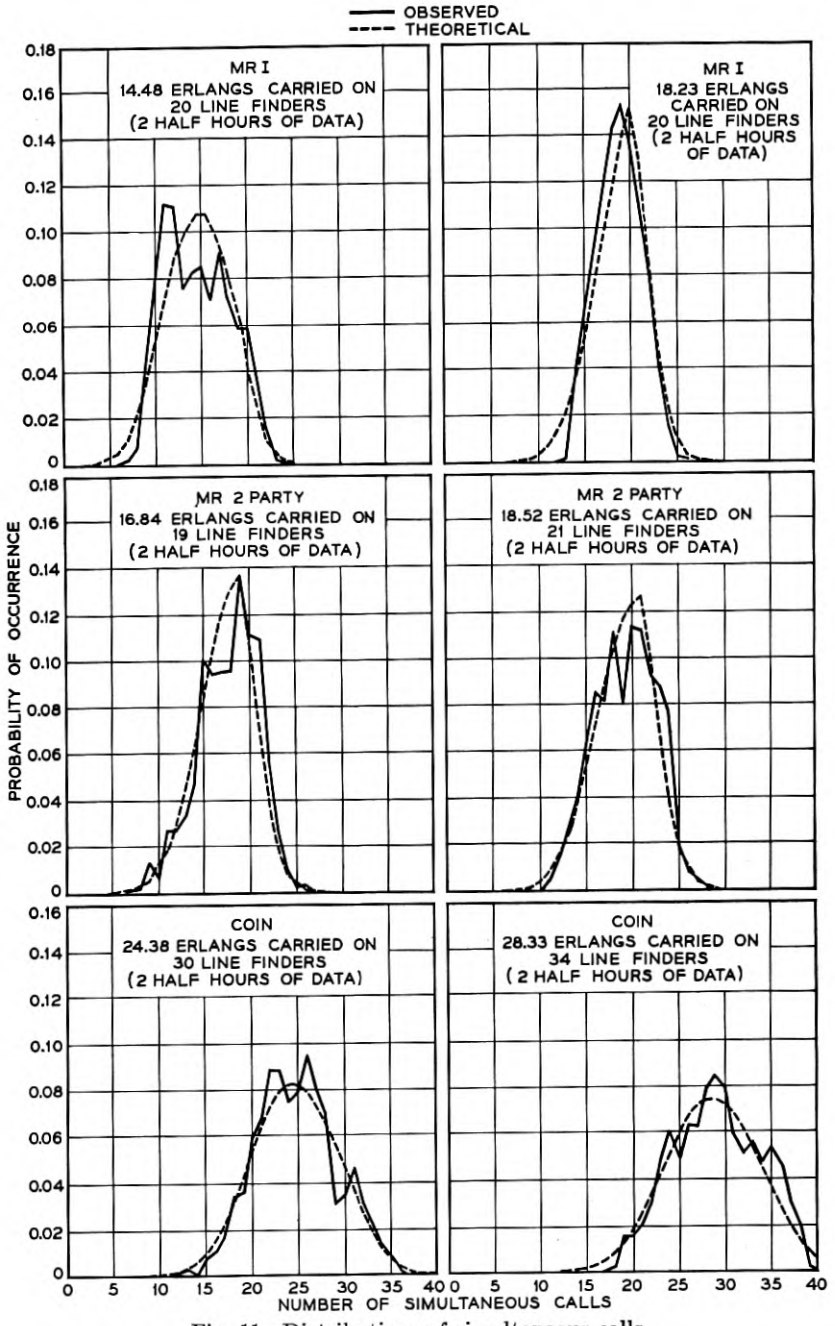


Fig. 11—Distributions of simultaneous calls.

The observer was provided with a means for introducing either no delay or one of four values of delay 2, 5, 10 or 15 seconds into the sender dial tone circuit. The observer took 50 observations using a particular value of dial tone delay and then shifted to another so that no particular value of delay would become evident to the customers during an afternoon's test. Each group of 50 observations comprised a mixture of message rate, coin and flat rate calls in the approximate proportions of 13 to 6 to 1, representing the respective volumes of traffic from these classes of service during the afternoon periods. It was not possible to distinguish PBX lines or two-party lines from the bulk of the message rate data nor PBX lines in the flat rate data, although to a limited extent the observer could identify PBX dialing by the generally faster pulsing. The coin data represent both public and semi-public customers.

Fig. 12 is a diagram for explaining the results shown on Figs. 13, 14 and 15 for the message rate, flat rate and coin classes of service, respectively as obtained with the sender monitoring circuit. Fig. 12 was obtained by the application of fitting curves to those message rate data of Fig. 13 for which a dial tone delay of five seconds was introduced by the observer. In the interval from $t = 0$ to $t = 5$ seconds, three curves A, B and C represent the per cent of subscribers still waiting at time t for dial tone. Curve A and its extension beyond $t = 5$ seconds represents the action of subscribers who would dial their calls before tone if dial tone were delayed indefinitely. Curve B and its extension beyond $t = 5$

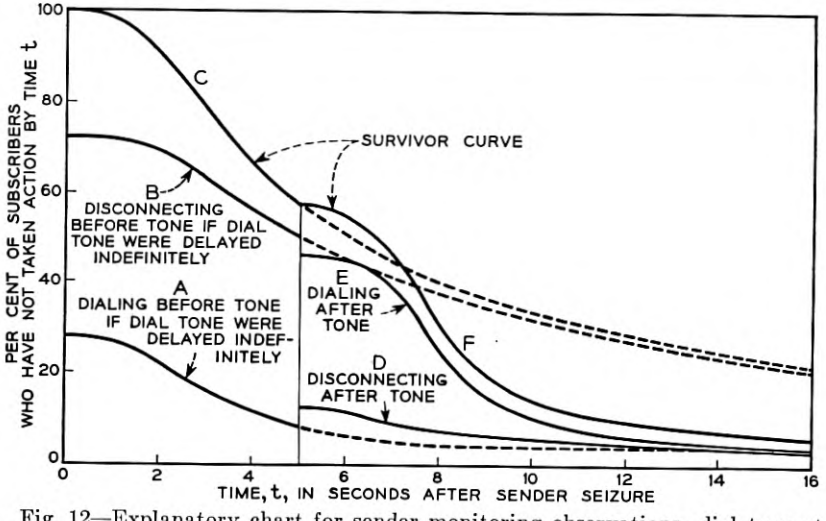


Fig. 12—Explanatory chart for sender monitoring observations; dial tone at $T = 5$ seconds.

seconds represents actions of subscribers who would disconnect if dial tone were delayed indefinitely. Curve C and its extension is the sum of the other two. Curve D in the region beyond $t = 5$ seconds represents the actions of subscribers who disconnect after tone, curve E represents the actions of subscribers who will dial their calls after tone and curve F represents the sum of the lower curves. Of interest is the fact that for an interval of about two seconds following dial tone (at $t = 5$ seconds), the observed total survivor curve F lies above the extended portion of

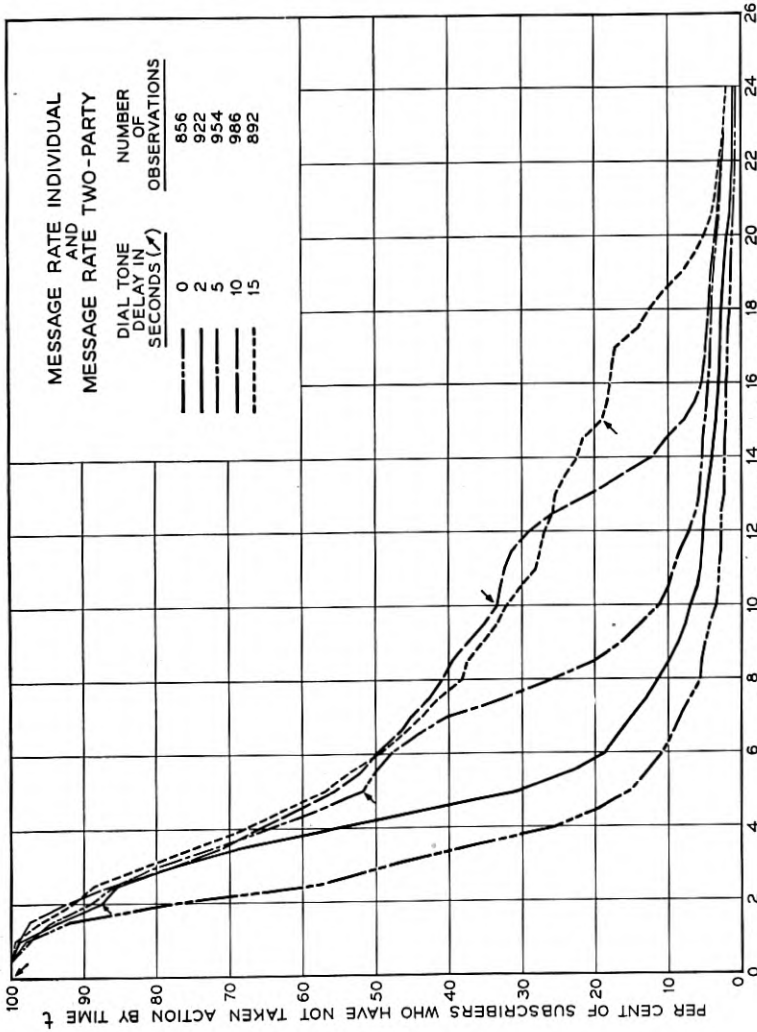


Fig. 13—Results of sender monitoring observations.

the hypothetical survivor curve C for infinitely delayed dial tone. This indicates that most of the subscribers who would have abandoned their attempts during this interval abruptly changed their minds and then consumed a noticeable interval of time after hearing tone before starting to dial. Thus, as might be anticipated, the subscribers exhibit a reaction time.

Fig. 13 shows the results in terms of survivor curves that were observed for the message rate class of service. Five sets of curves are

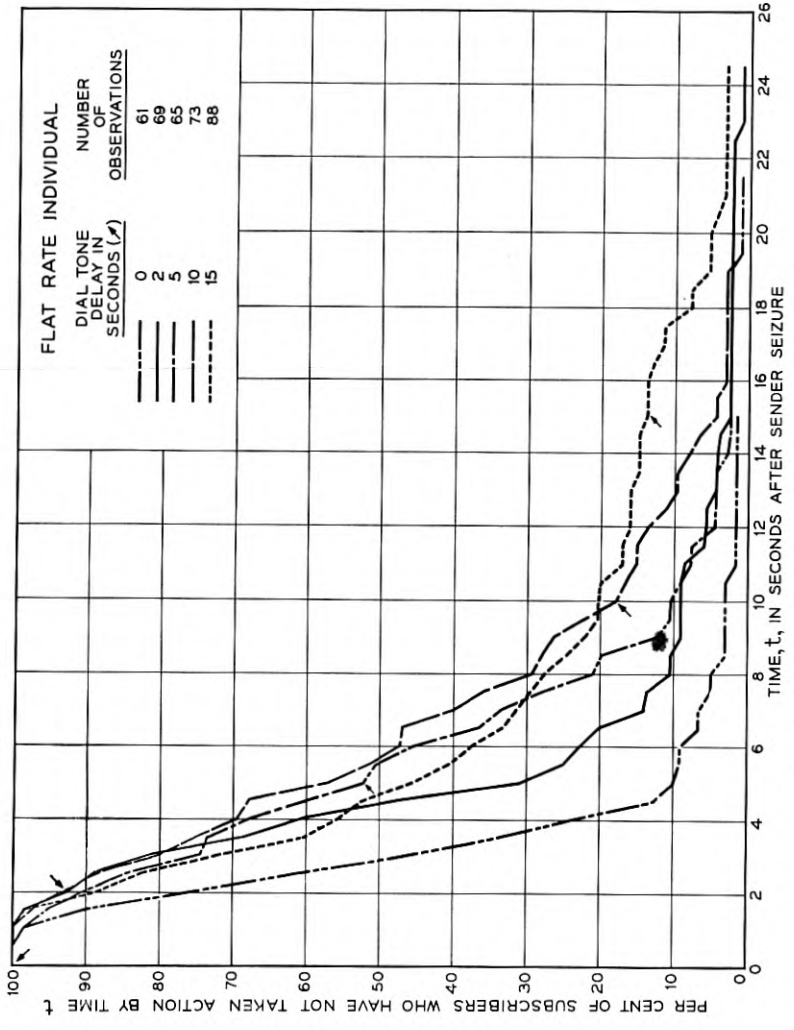


Fig. 14—Results of sender monitoring observations.

presented, namely for 0, 2, 5, 10 and 15 seconds of dial tone delays from the instant of sender seizure. It should be noted that the general contour of the various curves up to the receipt of dial tone and when extended beyond gives an estimate of the survivor curve for dial tone delayed indefinitely. Fig. 14 shows the results for the flat rate class of service (the data here are relatively meager), and Fig. 15 shows the results for coin customers.

Fig. 16 indicates for the three classes of service the progressive changes,

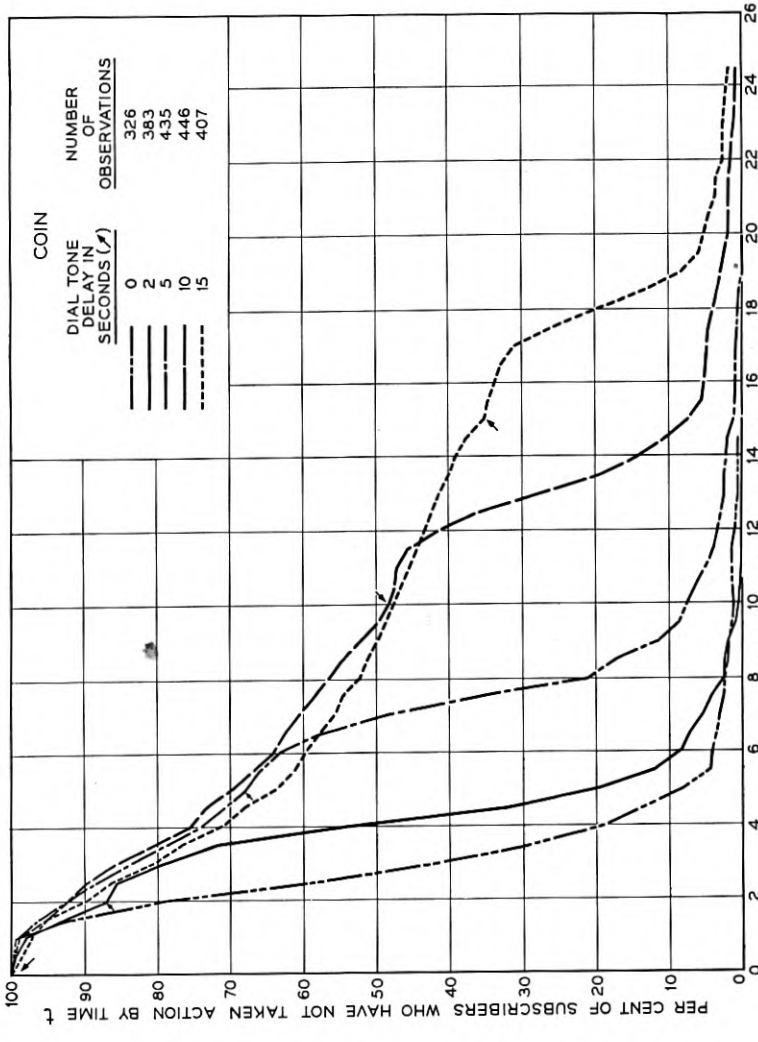


Fig. 15—Results of sender monitoring observations.

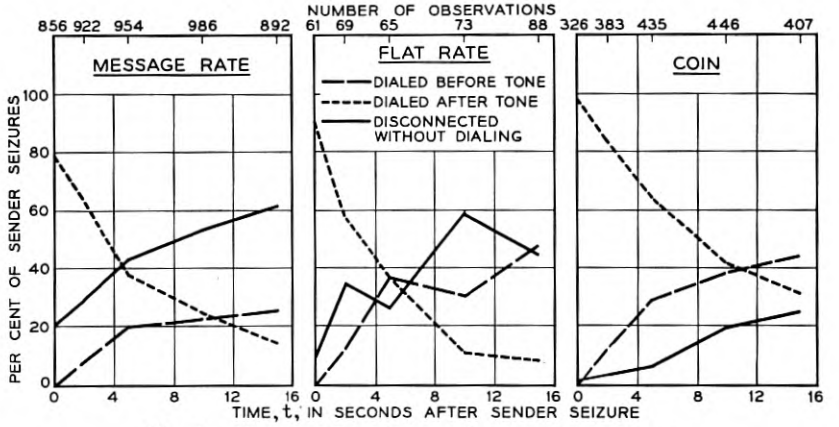


Fig. 16—Results of sender monitoring observations.

with increasing dial tone delay, of the per cent sender seizures resulting in dialing before tone, in dialing after tone and in disconnections for the five dial tone delay intervals studied.

Some general comments concerning Figs. 13 to 16 may be made.

1. There is a striking contrast between the message rate and coin classes of service. This may be due to the immediate financial stake that a coin customer has in his call. He is reluctant to disconnect before dial tone.

2. The results for the message rate and flat rate classes of service appear to be similar at the shorter dial tone delays; at the longer delays a higher proportion of flat rate subscribers have already dialed before dial tone. This apparent discrepancy may be due to the relatively small number of flat rate observations. Flat rate service in Sterling-3 and Main-2 was principally for professional people, such as doctors and nurses, who were thought to be more demanding than ordinary subscribers for prompt dial tone service. These results therefore should not be considered characteristic of flat rate customers generally.

3. Detailed analysis of the data (results not presented in this article) indicated that the distributions of time to first pulse for subscribers not observing tone, and for those waiting for tone, are quite similar for the three classes of service.

4. Only the unsmoothed raw data have been shown on Figs. 13 to 15 since certain inadequacies were detected in the observations. These were due to observer reactions and reading errors discovered as a result of comparing preliminary practice sender monitor test results with a simultaneous record obtained by the 100 pen tape recorder. The overall

effect is that the data obtained by the observer are generally displaced outward along the time axis by about 0.8 second.

5. The message rate data were for individual, PBX and two-party subscribers and the flat rate data were for individual and PBX subscribers. Furthermore, certain of the flat rate subscribers had auxiliary message rate service. It seems likely that different characteristics would be obtained for the individual, PBX and two-party subscribers since there appear to be reasons for expecting significant differences in their dialing habits. The PBX operator is in a position to "shop" for telephone service. If she fails to get dial tone on one outgoing line, she can try any other free line. This can also be done by subscribers with multiline service. This "shopping" for service tends to produce a large volume of disconnections when dial tone is slow. The individual and the two-party subscribers cannot do this and hence they can be expected to show fewer disconnections.

6. The results are in terms of intervals of time from the instant the sender is seized. It would, of course, be preferable to have these results in terms of time from receiver off hook. Since on the average the sender is seized in a time interval of about the same magnitude as that of the reaction time of the observer, Figs. 13 to 15 can be read approximately correctly when the abscissas are redesignated "time in seconds from receiver off hook." The foregoing results have been presented to furnish an increased understanding of subscriber dialing habits. In the next section additional results based on individual line records taken on the tapes are presented.

SUBSCRIBER DIALING HABITS OBSERVED BY INDIVIDUAL LINE RECORDS

As indicated in the previous section, the results obtained by means of the sender monitor tests were subject to certain shortcomings, hence data taken on individual lines with the 100-pen recorder have been analyzed to augment the information concerning the dialing habits of subscribers.

As noted heretofore, several of the pens on the 100-pen recorder were available for taking observations on subscribers lines. Two pens were used per subscriber line, one recorded the operation of the subscriber's line relay while the other pen marked whenever the subscriber's line was busy. On an originating call both pens started marking when the subscriber initiated a call. When a line finder was obtained, the line relay pen ceased marking and it was presumed that the subscriber obtained dial tone at that instant. Dialing, hang-up and flashing by a

subscriber after receipt of dial tone are noted by breaks in the markings of the line busy pen. Dialing, hang-up or flashing before dial tone are noted by simultaneous breaks in the markings of both pens. Various intervals can be measured and the call attempts classified accordingly.

Observations were obtained in the foregoing manner during the course of the Sterling-3 line finder tests on the following numbers of subscriber lines:

Message rate—residential	87
—business	23
—PBX	12
—two party	32
Flat rate individual	7
Coin	21
	182

The observed data were classified for each of the above six types of subscribers in terms of the following categories:

1. Time to subscriber action before receipt of dial tone.
   a. Time from receiver off hook[9] to first digit dialed by subscriber.
   b. Time from receiver off hook to disconnecting action by the subscriber.
2. Time from receiver off hook to receipt of dial tone.
3. Time to subscriber action after receipt of dial tone.
   a. Time from receipt of dial tone to first digit dialed by subscriber.
   b. Time from receipt of dial tone to disconnecting action by the subscriber.

Because the data developed in this section are compared with both the *j* factor analysis and the sender monitor test results and because the treatment of subscriber dialing habits before dial tone for each of these items is different, categories 1a and 1b are analyzed in two ways. In the *j* factor analysis, all actions of a subscriber prior to dial tone, except a disconnect, were considered to be one continuing demand for service. Thus for the first analysis, cases of dialing, flashing and short disconnections before tone lasting less than three seconds were ignored. In the sender monitor tests the only items considered were the time to the first digit dialed by a subscriber and the time, if no dialing occurred-to the release of the sender by the subscriber. Thus for the second analysis, cases of dialing before tone and flashing or disconnections before tone

[9] When a customer initiates a call, the line relay operates. For individual line and two-party subscribers this occurs when the subscriber takes his receiver off the hook. For coin customers this occurs when the customer has taken his receiver off the hook and made a proper deposit. For a PBX line this occurs when the PBX attendant has established a connection to an outside line. All of this is collectively termed "receiver off hook."

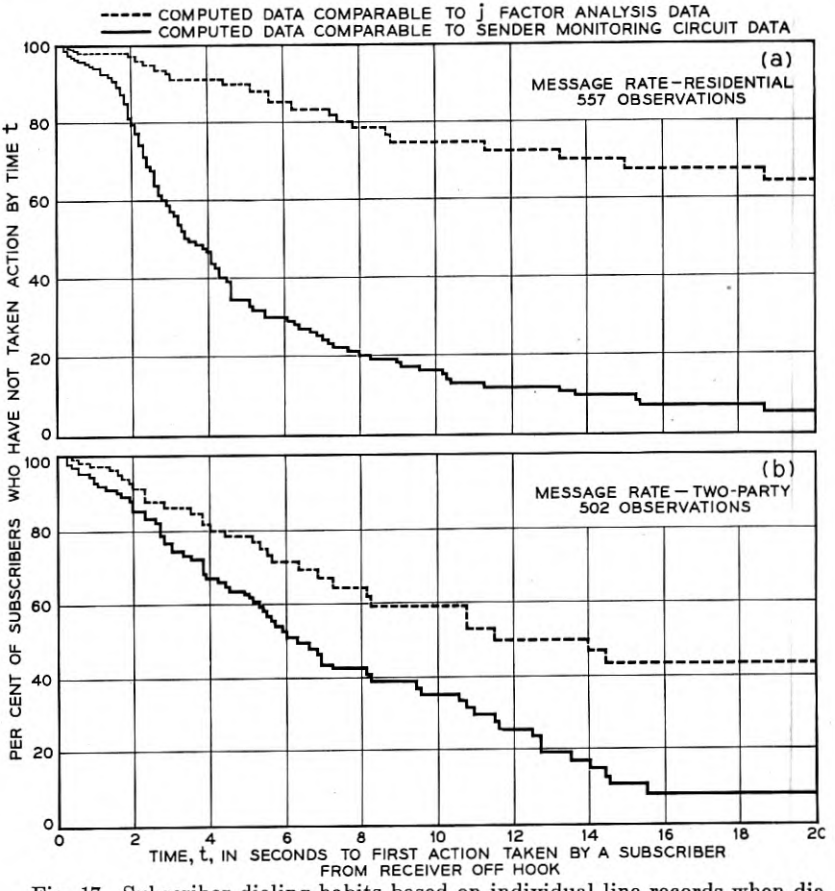


Fig. 17—Subscriber dialing habits based on individual line records when dia tone is delayed indefinitely.

that caused the sender to release were each counted as separate attempts. The results arrived at are shown as survivor curves on Figs. 17(a) to 19(b) for each of the six types of subscribers studied.

The survivor curves were developed by considering events during 0.1 second intervals. The number of cases of dialing before tone and the number of cases of disconnection before tone occurring during a 0.1 second interval were divided by the number of cases waiting for dial tone at the start of the interval. This ratio was considered to be a retirement rate. The complement of this rate gave a survival rate. By a progressive multiplication of survival rates from time, $t = 0$, the resulting survivor curves were obtained. Those cases receiving dial tone

during a particular 0.1 second interval are omitted from the number of cases waiting for dial tone at the start of the next interval.

In the development of the generalized trunking formula, the assumption was made that the waiting times of calls infinitely delayed have an exponential distribution. By assuming that the plots for the survivor curves in the development comparable to the j factor analysis are exponential distributions, it is possible by reading the value of t corresponding to 36.8 per cent of the subscribers still waiting for dial tone to obtain estimates of the values of H for the six types of subscribers. These estimates, most of which were obtained by extrapolation, are compared in

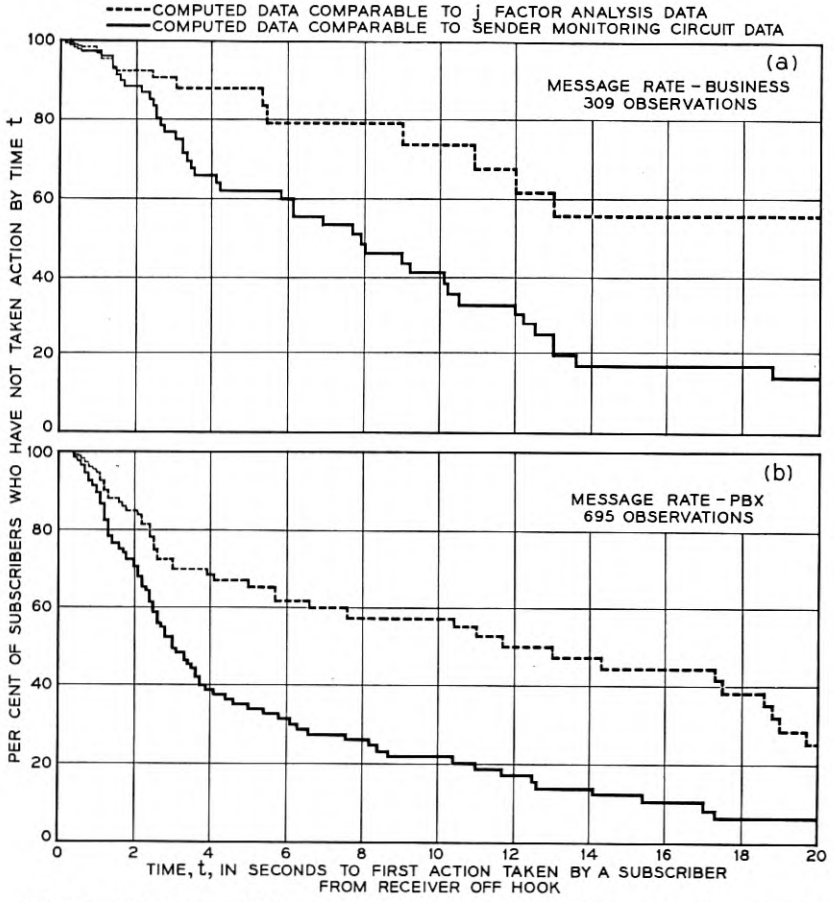


Fig. 18—Subscriber dialing habits based on individual line records when dial tone is delayed indefinitely.

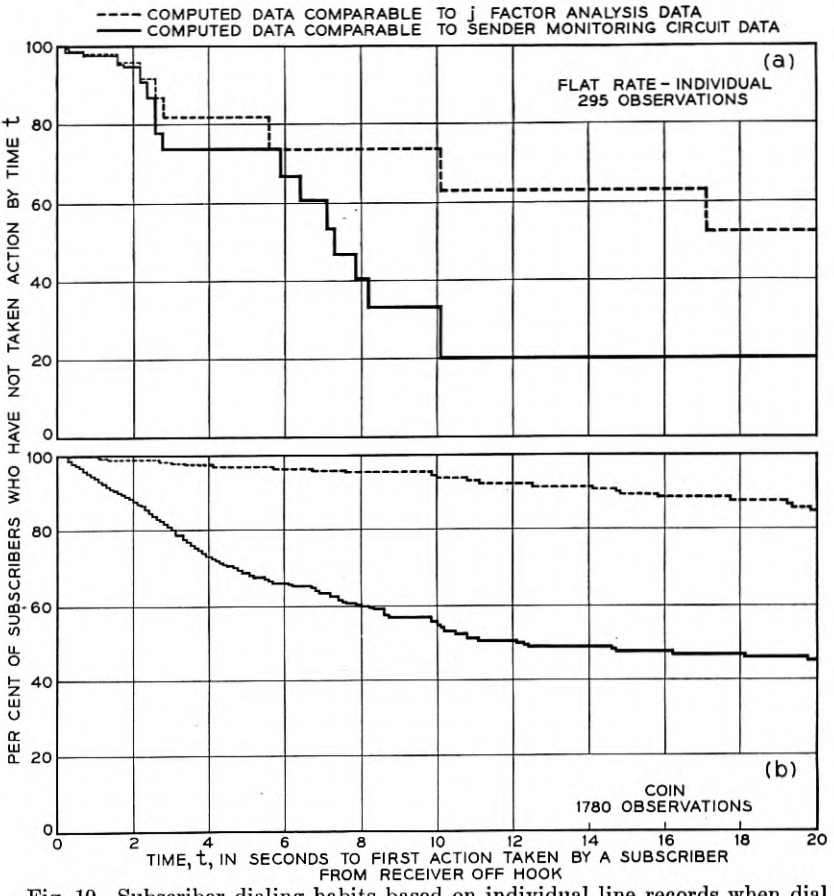


Fig. 19—Subscriber dialing habits based on individual line records when dial tone is delayed indefinitely.

Table IV with the values developed earlier in connection with the j factor analysis.

These results are not considered to be inconsistent since the tails of the survivor curves were constructed from very meager data. Conclusions based on Figs. 17(a) to 19(b) should therefore be regarded as having qualitative value only. The results principally indicate that the waiting-for-dial-tone characteristics of subscribers clearly vary with the different classes of service.

Comparisons between the lower survivor curves on Figs. 17(a) to 19(b) and the curves on Figs. 13 to 15 of the sender monitor tests are indicated by the percentages given in Table V of subscribers waiting for dial tone 5, 10 and 15 seconds from the time they requested service.

TABLE IV

	Estimated Values of H in Seconds	
	From figures 17(a) to 19(b)	From j factor analysis
Message rate—residential	45*	24
— business	33*	
— PBX	19	
— two-party	24*	42
Flat rate individual	32*	27
Coin	110*	74

* Rough extrapolated values

TABLE V

	Percentages of Subscribers Waiting for Dial Tone					
	Service Observation Figs. 17(a) to 19(b)			Sender Monitor Tests Figs. 13 to 15		
	5 secs	10 secs	15 secs	5 secs	10 secs	15 secs
Message rate						
Residential	34%	16%	10%	55%	33%	19%
Business	62	41	16			
PBX	35	22	13			
Two party	62	35	11			
Flat rate individual	74	34	20	55	19	14
Coin	68	55	47	67	48	35

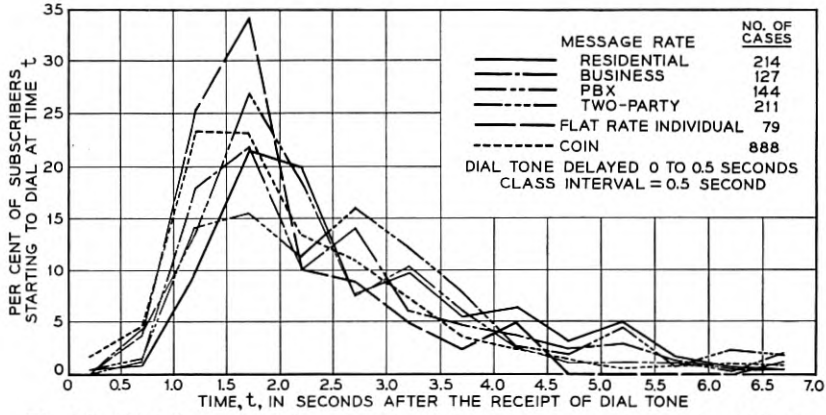


Fig. 20—Distributions of the start-to-dial times of subscribers who dial after the receipt of dial tone.

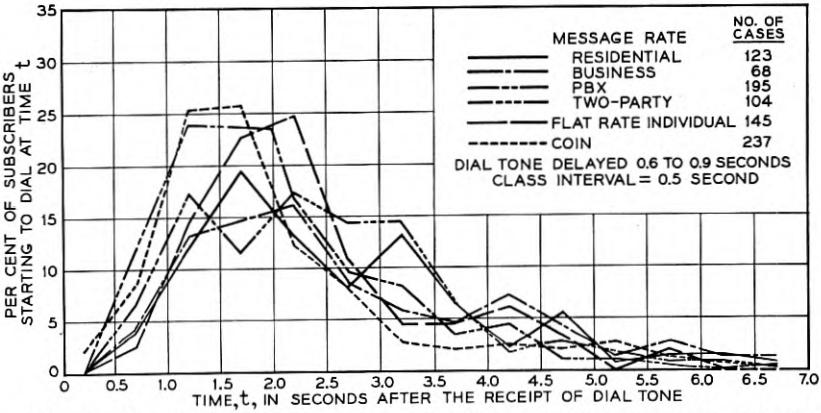


Fig. 21—Distributions of the start-to-dial times of subscribers who dial after the receipt of dial tone.

These results agree reasonably well when it is recalled that parts of the individual line data were scanty and that the sender monitor tests included the effects of observer reactions.

Once dial tone is received, it appears that all types of subscribers tend to follow a uniform dialing pattern. Figs. 20 to 23 show for a class interval of 0.5 second the distributions of the per cent of subscribers who dial at time t for the six types of subscribers studied. Figs. 20, 21 and 22 show the distributions when dial tone is received from 0.0 to 0.5, 0.6 to 0.9 and 1.0 to 1.9 seconds after dial tone, respectively. These curves

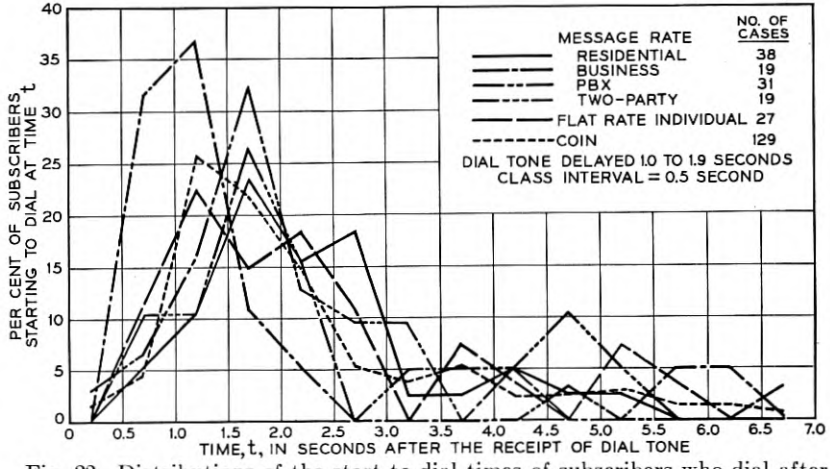


Fig. 22—Distributions of the start-to-dial times of subscribers who dial after the receipt of dial tone.

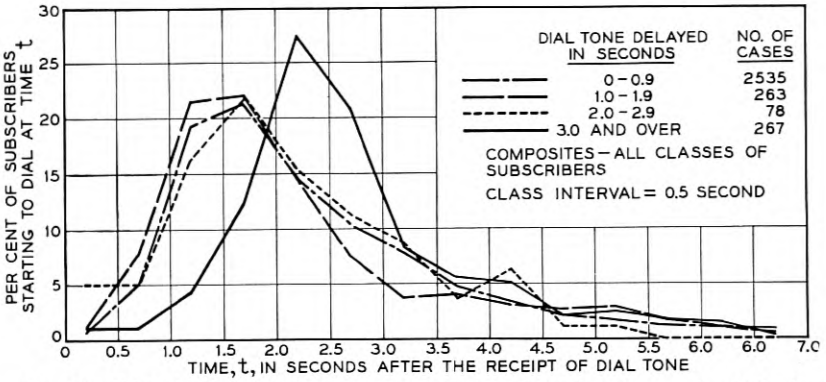


Fig. 23—Distributions of the start-to-dial times of subscribers who dial after the receipt of dial tone.

indicate that a strong similarity exists among the six types of subscribers with regard to their dialing patterns once dial tone is received.

Fig. 23 shows composite dialing distributions of the six types of subscribers for four dial tone delay intervals. For dial tone delays less than 3 seconds the dialing patterns are seen to be similar. Beyond 3 seconds delay the start-to-dial curve shifts outward, although the data are too scattered to indicate closely where the movement begins.

CONCLUSION

The foregoing report of the results of the tests conducted at the Sterling-3 central office indicates that the dialing habits of subscribers waiting for dial tone can be observed and analyzed to develop so-called j factors for use in a more general trunking formula than has been employed until recently in the Bell System. The report also presents descriptive data regarding the patterns subscribers follow when waiting for dial tone.

ACKNOWLEDGEMENT

The writers gratefully acknowledge the help given by the personnel of the Long Island Area of the New York Telephone Company, by their associates and others in the various phases of the study and particularly by C. F. Bischoff who had a major part in conducting the tests.

Selective Fading of Microwaves

BY A. B. CRAWFORD AND W. C. JAKES, JR.

(Manuscript received October 25, 1951)

The results of an extended survey of microwave propagation over two line-of-sight paths in New Jersey are described. Angle-of-arrival measurements at 1.25-cm wavelength and selective fading observations in a 450-mc frequency band centered at 3950-mc show that the severe fading can be explained in terms of multiple-path transmission. A computer of the analogue type was built to simulate the more complicated selective fading patterns.

INTRODUCTION

During the past few years, studies of microwave propagation have been made by the Radio Research group at the Holmdel Laboratory over two paths located in eastern New Jersey. Both of these are line-of-sight paths which might be considered to be typical links in a cross-country microwave radio relay circuit.

In conducting these studies, the usual continuous recordings of signal levels were made but the greater interest was centered in special experiments designed to reveal more of the processes which can cause fading. The most relevant information has been obtained by exploring the incident wavefronts with a narrow-beam scanning antenna (angle-of-arrival studies) and, more recently, by observing the transmission characteristics of the paths by means of a frequency-sweep technique and also by the use of very short pulses.

Some results of angle-of-arrival observations have been reported previously[1] and a companion paper describes the transmission tests conducted with very short pulses.[2] The present paper describes some of the observed mechanisms associated with fading, presents typical data obtained with the narrow-beam scanning antenna and gives examples of the frequency-sweep observations, illustrating the frequency selective

[1] W. M. Sharpless, "Measurement of the Angle of Arrival of Microwaves," *Proc. I.R.E.*, **34,** Nov. 1946, pp. 837–845. A. B. Crawford and W. M. Sharpless, "Further Observations of the Angle of Arrival of Microwaves," *Proc. I.R.E.*, **34,** Nov. 1946, pp. 845–848. H. T. Friis, "Microwave Repeater Research," *Bell System Tech. J.*, **27,** Part I, "Propagation Studies" by A. B. Crawford, Apr. 1948, pp. 183–246.

[2] O. E. DeLange, "Propagation Studies at Microwave Frequencies by Means of Very Short Pulses," *Bell System Tech. J.*, **31,** Jan. 1952, pp. 91–193.

nature of the fading. Some data derived from the continuous recordings of signal levels are presented in an appendix.

The angle-of-arrival observations were made at a frequency of 24,000 megacycles. The frequency-sweep experiment and the recordings of signal levels were made in the 3700 to 4200 megacycle frequency band as were the short pulse observations described in the companion paper.

GENERAL DISCUSSION OF PROPAGATION PHENOMENA

The map of Fig. 1 shows the location of the experimental transmission paths. The path between Crawford Hill and Southard Hill is 17 miles long and clears the intervening terrain by 65 feet, approximately one Fresnel zone at a frequency of 4000 megacycles. The other path, between Crawford Hill and a 100-foot tower on the Murray Hill Laboratory property, is 22.8 miles long and has clearance of 280 feet. Fig. 2 shows the profiles of these two paths.

The general characteristics of over-land microwave transmission are well known and need be reviewed only briefly. During the daytime hours, when the lower atmosphere is thoroughly mixed by rising convection currents and winds, the signals are normally stable and are near

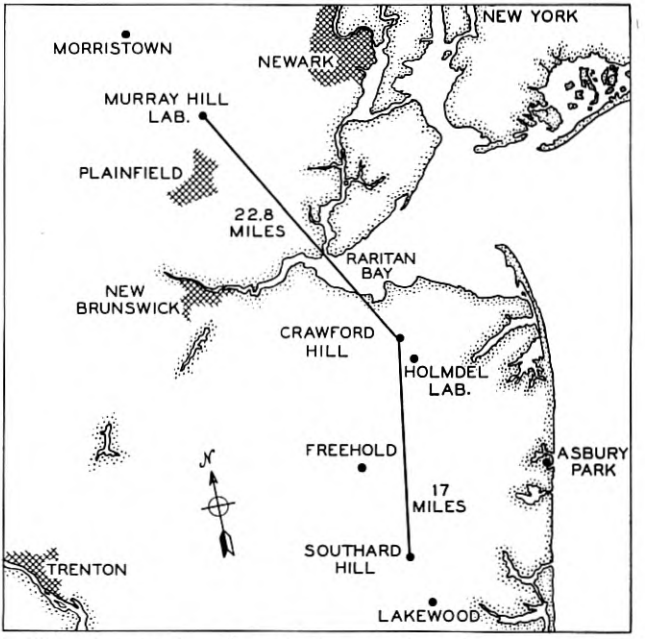


Fig. 1—Map showing location of the experimental transmission path.

the free space levels. Also during the winter months, when the humidity content of the atmosphere is low, signal variations are usually very small. However, on clear summer nights with little or no wind, non-uniform distributions of temperature and humidity can create steep dielectric constant gradients in the lower atmosphere, thus causing anomalous propagation and fading.

When fading occurred on our experimental transmission paths, an alarm circuit connected into the continuously recording equipment was arranged to operate when the signal level dropped below a predetermined value. This enabled observers to be present during severe fading periods

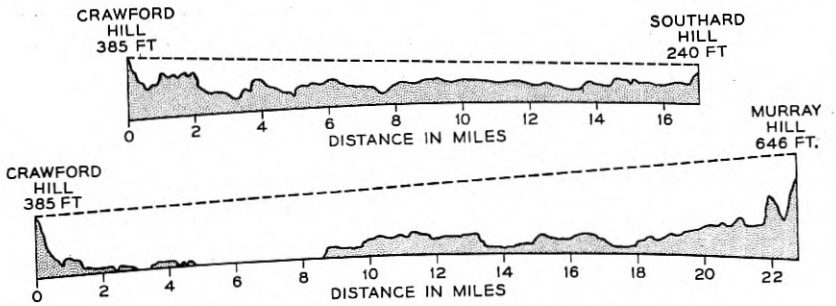


Fig. 2—Profiles of the transmission paths.

and to seek, by means of the special experiments, to determine the causes of the fading.

Although it has not been possible to provide satisfactory explanations for all of the observed fading phenomena, much of the fading (occasions when the signals are depressed to levels 15 to 20 decibels or more below the normal daytime value) can be explained qualitatively in terms of simple ray pictures. Fig. 3 is intended to illustrate some of the observed fading phenomena. The case of multiple path transmission, the most common cause of fading on either transmission path, is shown in Fig. 3(a). Two, three and sometimes more signal components are found to arrive at various angles in the vertical plane, usually above the line of sight. Wave interference among these components produces fading, the severity of which depends upon the relative amplitudes and delays of the components. At these times, different frequencies fade differently and the signals received on two vertically spaced antennas also fade differently. The use of either frequency or space diversity would be effective in this type of fading.

A relatively rare type of fading, observed only on the Murray Hill path, is believed to be caused by the mechanism illustrated in Fig. 3(b).

Here a reflecting layer is situated between the heights of the transmitter and receiver. The signal then suffers attenuation due to reflection of part of the energy from the direct path. Widely separated frequencies are affected in like fashion and the outputs of antennas spaced for diversity reception tend to be in agreement although the fine structure fading is usually different.

On neither of the experimental transmission paths is there a rgeular ground-reflected component of any consequence. Due to the roughness of the ground and the presence of vegetation, the effective reflection coefficient is of the order of 0.2 for either path. Ground reflections thus play

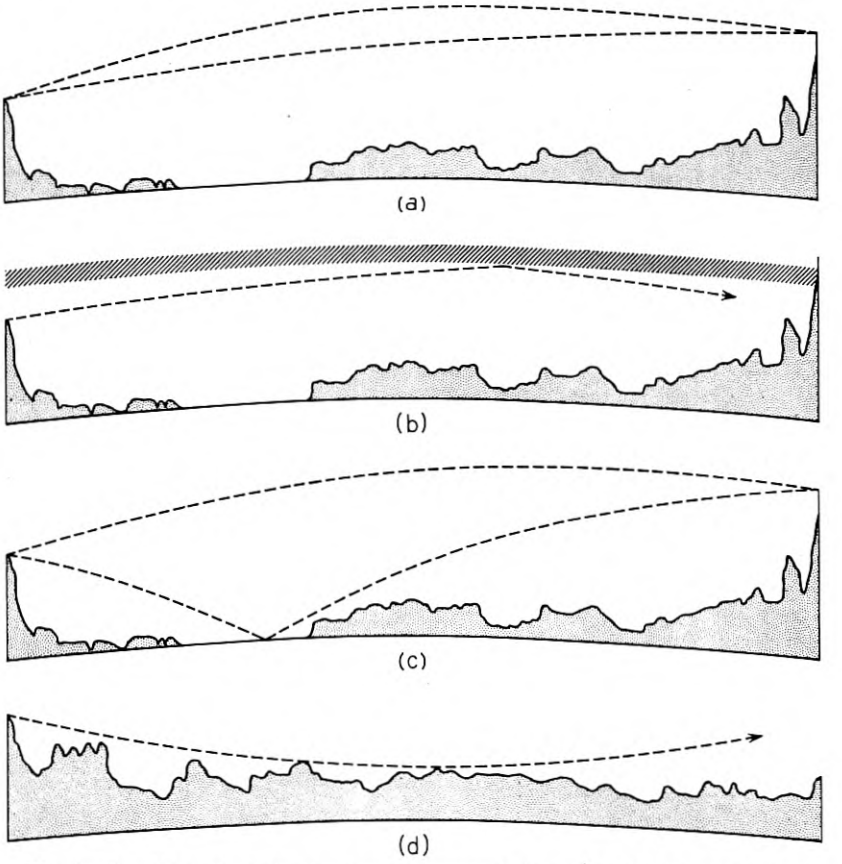


Fig. 3—Possible ray paths involved in severe fading. (a) Multiple path transmission. (b) Attenuation by reflection from an elevated layer. (c) Abnormal water reflection on the Murray Hill-Crawford Hill path. (d) Substandard conditions on the Southard Hill-Crawford Hill path.

no significant part in the fading picture with the exception of the situation illustrated in Fig. 3(c). Occasionally on the Murray Hill path, conditions of atmospheric refraction are such that a strong signal component is received by virtue of reflection from the water surface of Raritan Bay. Under normal conditions, the geometry of the path does not permit such a reflection.

Normally the dielectric constant of the atmosphere decreases with height above ground so that the ray path usually has a curvature in the same direction as the earth curvature. However, it is possible for the dielectric constant of the atmosphere to increase with height above ground (sub-standard conditions) so that the ray path has a curvature opposite that of the earth. This results in the condition illustrated in Fig. 3(d) where the limiting or tangent ray does not reach the receiver and only a weak signal is received by virtue of diffraction. Widely separated frequencies and vertically spaced antennas are affected alike as regards the average signal level but not the fine structure fading. This effect has been observed only on the Southard Hill-Crawford Hill path which has small clearance to begin with. It has been observed on several nights in late summer or early autumn after a radiation type ground fog has formed in the late evening and usually persists until the fog is dispelled by winds or by the morning sun.

There are, of course, times when the transmission conditions are considerably more complicated than those described above. Some of these apparently are due to a combination of the situations illustrated in Fig. 3 while others may be the result of an atmospheric focussing or trapping phenomenon. In addition to the various phenomena just described, which, fortunately, occur rather infrequently, there are numberous occasions when the signal varies plus and minus a few decibels relative to the free space level. It has not been possible actually to demonstrate the mechanism responsible but it seems most likely that these smaller variations are due to non-linear dielectric constant gradients which give the atmosphere the properties of a convergent or divergent lens.

An important result of the observations made to date is the conviction that the severe fades, signal excursions to levels 30 decibels or more below the free space field, were all caused by wave interference. It appears that, as the average signal level is depressed by any mechanism, it becomes more and more vulnerable to the effects of extra signal components of small amplitude that often may be present but go unnoticed when the signal is near normal levels. Thus, while the average signal level during the conditions illustrated in Figs. 3(b) and 3(d) may be no more than 15 to 20 decibels below the normal daytime level, there

is usually superimposed a fine structure fading in which short duration fades to levels as much as 45 decibels below free space have been observed. For this reason it is desirable to avoid paths having small clearance over intervening terrain and also paths which have a permanent ground reflection of sufficient magnitude to depress the signal to critical levels when, due to variable atmospheric refraction, the direct and reflected components are in phase opposition.

The following sections describe the angle-of-arrival and frequency-sweep experiments on which much of the preceding discussion was based.

ANGLE-OF-ARRIVAL OBSERVATIONS

A photograph of the Crawford Hill receiving site is shown in Fig. 4. The building housing the receiving equipment and the associated antennas are mounted on a framework which can be rotated on a concrete track, permitting investigation of the transmission characteristics of either path. The parabolic antennas on the tower are used for continuous recording of 4195 megacycle signals. The long object at the left of the

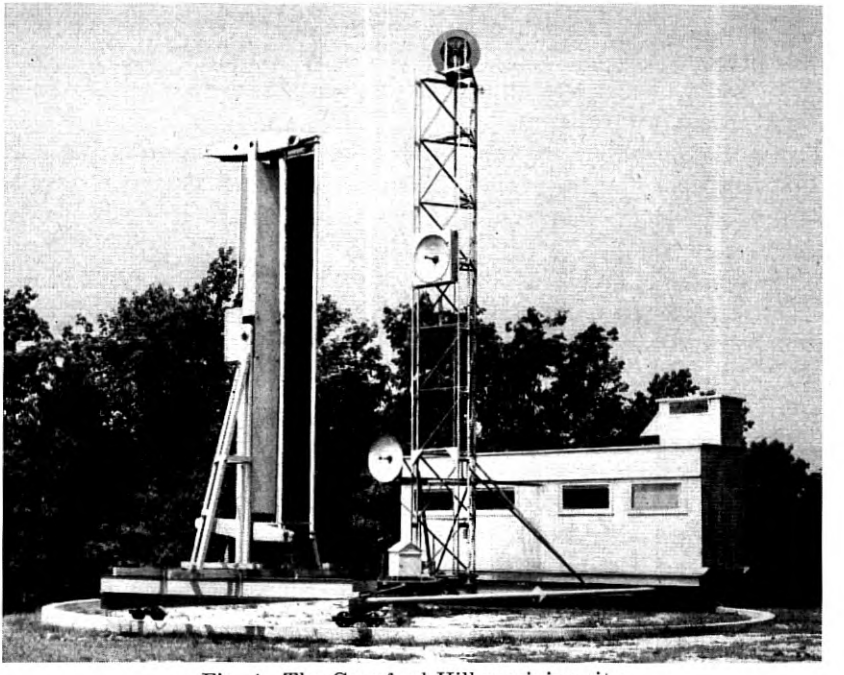


Fig. 4—The Crawford Hill receiving site.

picture is the metal-lens antenna used for making the angle-of-arrival observations. Its half-power beamwidth is 0.12 degree at the operating frequency of 24,000 mc. The focal length of the lens is 48 feet and its feed is located in the little cupola on top of the building. The feed is held fixed, while the lens is moved vertically by a motor-driven mechanism; thus the antenna beam also moves vertically. The antenna scans a total angle of two degrees in ten seconds. It is fed by a 24,000-mc radar set which is gated to receive only the pulses reflected from a corner reflector located at the distant terminal of the transmission path. The spot on the radar cathode ray tube moves vertically in synchronism with the scanning antenna, and the horizontal deflection is proportional to the amplitude of the pulse received from the corner reflector. The display thus shows amplitude of the various incoming signal components as a function of their angles of arrival.

The antenna installation on Southard Hill is shown in Fig. 5. At the left is the transmitting paraboloid for the 4195-mc continuous wave transmitter, the radar corner reflector is in the center, and on the right is the horn-reflector antenna used in the frequency-sweep experiments described below. Similar equipment is located at the Murray Hill terminus. The corner reflector is 5.5 feet on a side, and at 24,000-mc has sufficient gain to override reflections from other nearby objects, and thus becomes easily identifiable on the radar screen.

The radar oscilloscope for typical propagation conditions is shown in Fig. 6. These pictures were obtained by leaving the camera shutter open during the ten-second interval required for the antenna beam to scan through the angular range of 2°. All of these representative photographs were taken on the Murray Hill-Crawford Hill path although similar results were obtained on the Southard Hill-Crawford Hill path with the exception of Fig. 6(f). The normal daytime transmission is shown in Fig. 6(a) to consist of a single path arriving at an angle of $-0.2°$ with respect to a fixed reference angle. The horizontal lines represent intervals of 0.1°, so that changes of 0.05° can be estimated. The other pictures in Fig. 6 were all taken during fading conditions.

Figs. 6(b) and 6(c) are good examples of the multiple-path condition shown in Fig. 3(a) in which the individual components are almost equal in amplitude and well separated in angle. In Fig. 6(b) there are two components arriving at angles of 0.1° and 0.6° above the normal line-of-sight while in 6(c) there are three components with angles of 0.05°, 0.35° and 0.7° above the normal angle. The position and amplitude of the signal components may change radically in a matter of minutes, and often there is no component that can be identified as the "normal" one.

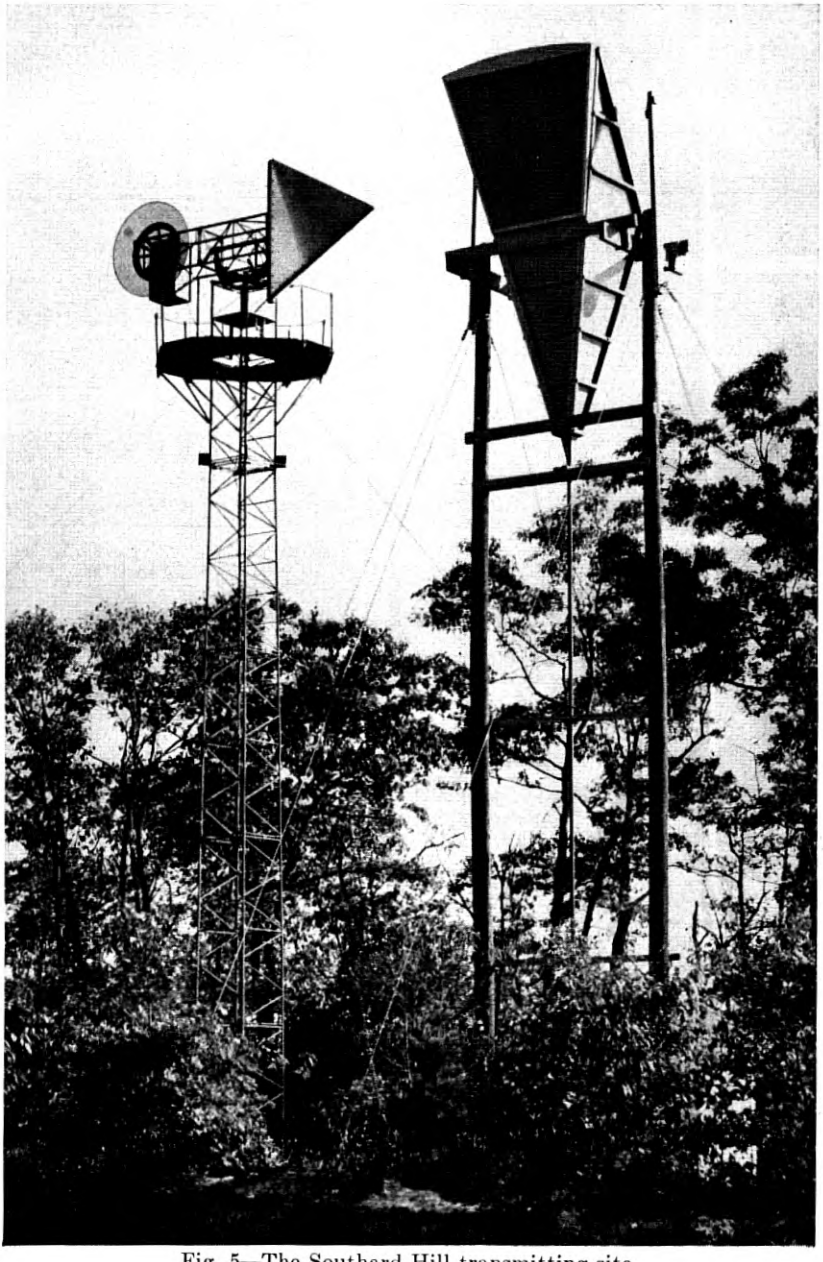


Fig. 5—The Southard Hill transmitting site.

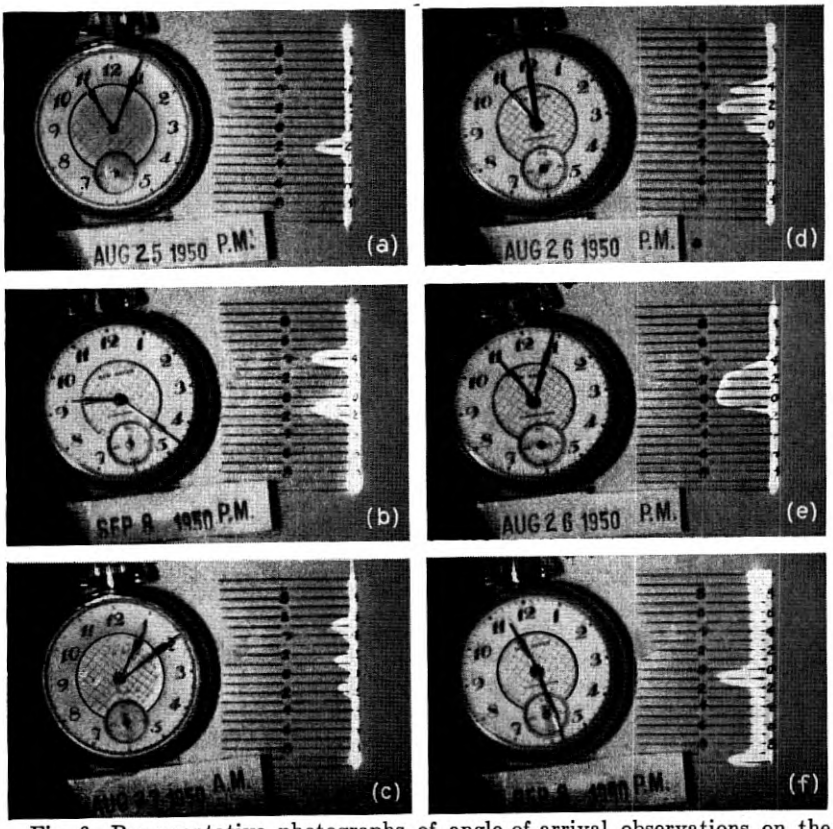


Fig. 6—Representative photographs of angle-of-arrival observations on the Murray Hill-Crawford Hill path. (a) Normal day. (b) Two elevated paths. Sept. 8, 1950; 9:23 p.m. (c) Three elevated paths. Aug. 27, 1950; 1:11 a.m. (d) Multiple paths. August 26, 1950; 11:00 p.m. (e) Wide angle "fill-in". Aug. 26, 1950; 11:04 p.m. (f) Abnormal water reflection. Sept. 8, 1950; 11:28 p.m.

During these multiple-path conditions, the recordings of the 4195-mc transmission generally show the broad maxima and sharp minima characteristic of wave interference.

Figure 6(d) shows a case in which the various paths are not completely separated while Fig. 6(e) (taken four minutes later) shows that energy is being received almost without variation over a vertical angle of 0.4°. This may represent a number of ray paths which would be separable by a narrower-beam antenna, or it may indicate a focussing or trapping phenomenon. Often when the type of transmission illustrated by 6(e) is present, the recorded 4195-mc signal may be as much as 12 to 15 decibels above the free space levels.

Fig. 6(f) illustrates the case of abnormal reflection from the water of Raritan Bay on the Murray Hill path as indicated in Fig. 3(c). Here the "normal" signal component is arriving at 0.1° above the line-of-sight while another component, almost equal in amplitude, is arriving at the very bottom of the scan, about 0.8° below the line-of-sight. It is quite probable that there have been times when this component was present but was outside the range of the scanning antenna.

The mechanisms discussed in connection with Fig. 3(b) and 3(d) cannot be demonstrated by photographs such as those just presented although the angle-of-arrival radar was instrumental in furnishing the clues to the phenomena. Due to the two-way attenuation of the radar-corner reflector technique, the signal at these times rapidly falls below the noise level of the receiver. For the same reason, it is not possible to detect the extra signal components of small amplitude which were postulated to account for the very deep fades sometimes observed under these transmission conditions.

FREQUENCY-SWEEP OBSERVATIONS

Since most of the fading is due to interference between waves which travel over different paths of, presumably, different lengths it was realized that the fading was likely to be frequency selective. Just how selective would depend on the relative lengths of the individual transmission paths. The usual methods for determining path length differences are to use short pulses, or to sweep the frequency. Since it was likely that the path-length differences would be measured in feet rather than yards, very short pulses or a wide frequency-sweep were required. An oscillator[3] was available whose frequency could be swept over the licensed band of 500 mc between 3700 mc and 4200 mc. The frequency-sweep experiment was set up on the Murray Hill-Crawford Hill path for the summer of 1949. The following summer, the milli-microsecond pulse transmission tests described in the companion paper were conducted over the same path. As might be expected, simultaneous observations showed good agreement between the two methods.

The frequency of the transmitter, located at Murray Hill, is swept over a 450-mc band centered at 3950 mc at a 60-cycle rate. At the receiver, a similar oscillator is used for the beating oscillator except that its frequency is swept linearly through the same frequency band in one

[3] This oscillator was developed by M. E. Hines and is described in his paper published in the *Bell System Technical Journal*, Vol. 29, Oct. 1950. It uses a 416A close-spaced triode in a wave-guide cavity. The frequency is changed by means of a plunger which is capacity-coupled to the plate of the tube and which is actuated by a modified loud speaker unit.

second. Since the intermediate frequency amplifier of the receiver is only 350 kc wide, (centered at 600 kc) narrow pulses are generated each time the frequency of the transmitter crosses the frequency to which the receiver is tuned. These intermediate frequency pulses are displayed vertically on a cathode ray tube. The horizontal trace is synchronized with the one-second sweep rate of the beating oscillator.

The normal daytime frequency-sweep pattern is shown in Fig. 7(a). The vertical scale is linear in amplitude and the horizontal scale is almost linear in frequency, with frequency decreasing from left to right. Visible at the extreme left is the signal used for continuous recording. Since there is only one transmission path involved, the amplitude of the received signal is nearly constant over the 450-mc band. If another signal were present which had travelled over a path of different length, the two signals would add when the frequency is such that the path length difference is an even multiple of half-wavelengths and subtract when the path length difference is an odd multiple of half-wavelengths. Simple calculation shows that if the path length difference is one foot, the frequencies at which the signals add and subtract are separated about 500 mc. Thus the limit of resolution for the frequency-sweep experiment is a little more than one foot.

Photographs taken on a night when the angle-of-arrival radar indicated two almost equal components separated about 0.4 degrees in angle are shown in Fig. 7(b). The time interval between the two pictures is 30 seconds, during which the minimum had shifted about 150 mc. The pictures can be interpreted as simple two-path transmission with an indicated path difference of about two feet and an amplitude ratio of 0.7 to 1. On this night the minimum shifted back and forth across the frequency band—sometimes slowly and sometimes rapidly. At times the position of the minimum might remain fixed but its depth would change.

Photographs taken on a night when there were abnormal reflections from the water of Raritan Bay are shown in Fig. 7(c). There are evidently two main components with path difference of about six feet, with a small third component causing the slight decrease in amplitude of the peaks from left to right. These pictures were taken 9 minutes apart, but this type of pattern was observed over a period of about three hours on this night.

Usually the frequency sweep patterns are considerably more complicated than the ones shown so far. Fig. 7(d) shows two photographs which indicate that at least three signal components and perhaps more were present. The time interval between the two pictures was about 30 seconds.

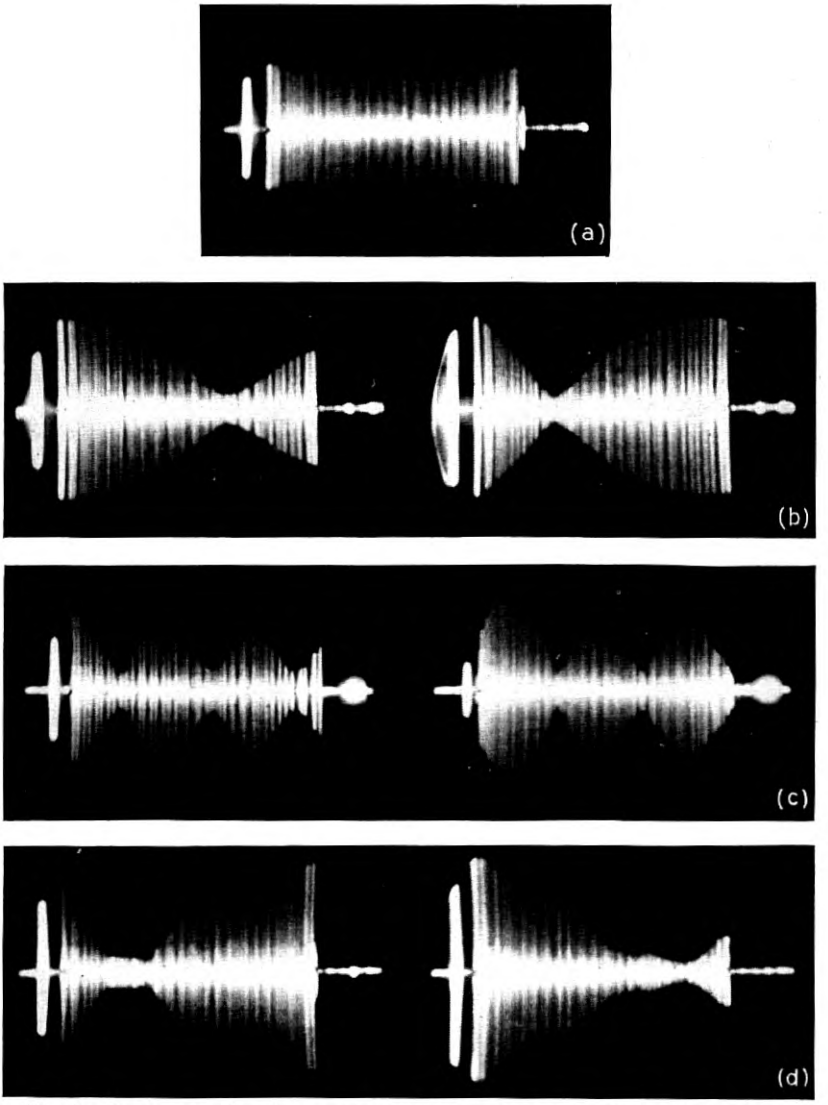


Fig. 7—Representative frequency-sweep patterns observed on the Murray Hill-Crawford Hill path. (Summer 1949.) (a) Normal day. (b) Two components with a path difference of two feet. (c) Two main components with a path difference of about six feet plus a small third component. (d) Multiple component pattern.

SYNTHESIS OF FREQUENCY-SWEEP PATTERNS

To aid in the interpretation of the complicated frequency sweep patterns, a computor of the analogue type was built. This apparatus combines four signal components, three of which are variable in delay and amplitude, and presents the result on a cathode ray tube in the same form as the actual frequency sweep patterns. Thus a particular pattern can be synthesized on the computor and the number of components, together with their path differences and relative amplitudes, read directly from the computor dials. This is accomplished by generating four 600-kc signals, three of which are phase modulated at 60 cycles per second. The total phase deviation and relative signal amplitude are variable. The four signals are then summed and displayed in vertical deflection on a cathode ray tube having a 60-cycle horizontal sweep.

The synthesis of the patterns of Figures 7(b) and 7(c) are shown in Fig. 8. The upper synthesized pattern is simply a combination of two components with relative amplitudes of 0.7 and 1 and a path difference of two feet. The lower pattern consists of the reference component with unity amplitude, a second component with an amplitude of 0.5 and a path difference of 5.7 feet, and a third component with an amplitude of 0.2 and path difference of 0.8 feet. The similarity between the actual and synthesized patterns is obvious.

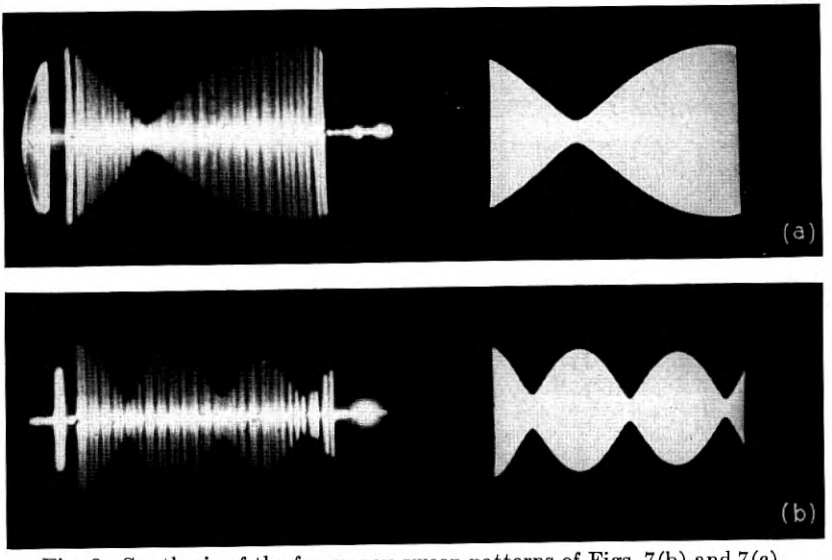


Fig. 8—Synthesis of the frequency-sweep patterns of Figs. 7(b) and 7(c).

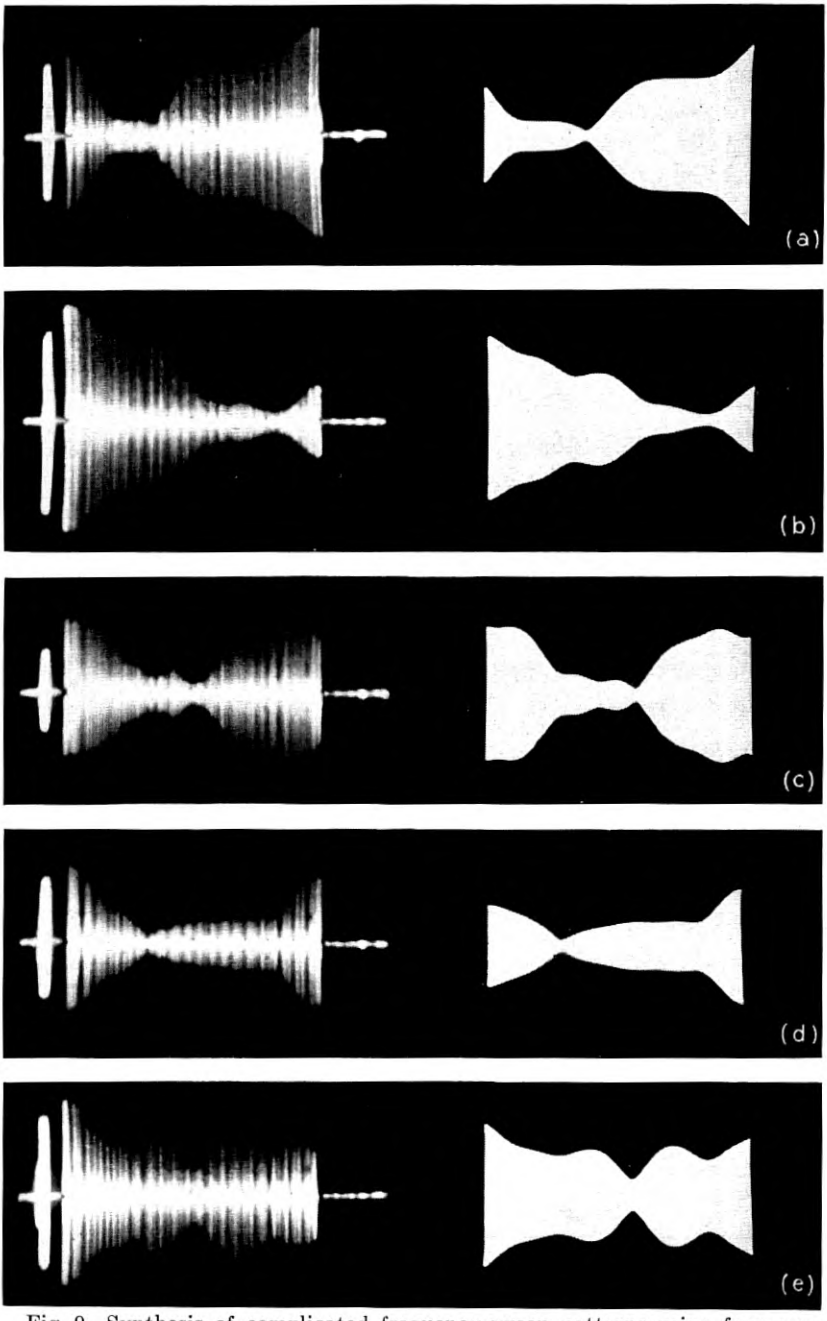


Fig. 9—Synthesis of complicated frequency-sweep patterns using four components. See Table I for values of the relative amplitudes and path differences.

Examples of attempts to synthesize some more complicated frequency sweep patterns, taken on the night of August 2, 1949, are shown in Fig. 9. Four components were required in each case, with the path differences and delays being summarized in Table I. Although the pictures all appear very different, in general major changes were required only in component No. 2 to go from one pattern from the next, as Table I shows. All the remaining components had to be very carefully trimmed in both amplitude and delay to get good synthesis (especially in the case of Fig. 9(d), but these changes were relatively small.

CONCLUDING REMARKS

The special experiments just described have led to the conclusion, expressed earlier, that the severe fading observed on the two test paths is the result of multiple-path transmission in which several components may be involved. These components may arrive at the receiver at various angles up to three quarters of a degree above the normal daytime angle-of-arrival and, in the case of abnormal water reflection on the Murray Hill path, as much as 0.8 degree below the normal angle. The path differences among these components may vary from a fraction of a foot to about ten feet. The long-delay components are usually small in amplitude.

In all cases where observations were made during periods of exceptionally high signal levels, say 10 to 15 decibels above free space level, the frequency-sweep patterns were substantially flat, suggesting a focussing or trapping phenomenon. The frequency-sweep patterns were also flat on those nights when the signal excursions were only a few decibels above and below the normal daytime level. However, the severe fades

TABLE I

		Fig. 9 (a)	Fig. 9(b)	Fig. 9(c)	Fig. 9(d)	Fig. 9(e)
Component No. 1 (Reference)	Amplitude	1	1	1	1	1
	Path diff. (ft.)	0	0	0	0	0
Component No. 2	Amplitude	0.9	1.2	0.7	1.1	0.4
	Path diff. (ft.)	1.1	0.5	1.7	0.5	1.1
Component No. 3	Amplitude	0.2	0.1	0.2	0.2	0.2
	Path diff. (ft.)	5.2	5.7	5.6	5.7	5.7
Component No. 4	Amplitude	0.05	0.1	0.15	0.1	0.1
	Path diff. (ft.)	9.2	9.2	11.0	9.3	8.7
Time		12:08½ AM	12:09 AM	12:18 AM	12:24 AM	12:25 AM

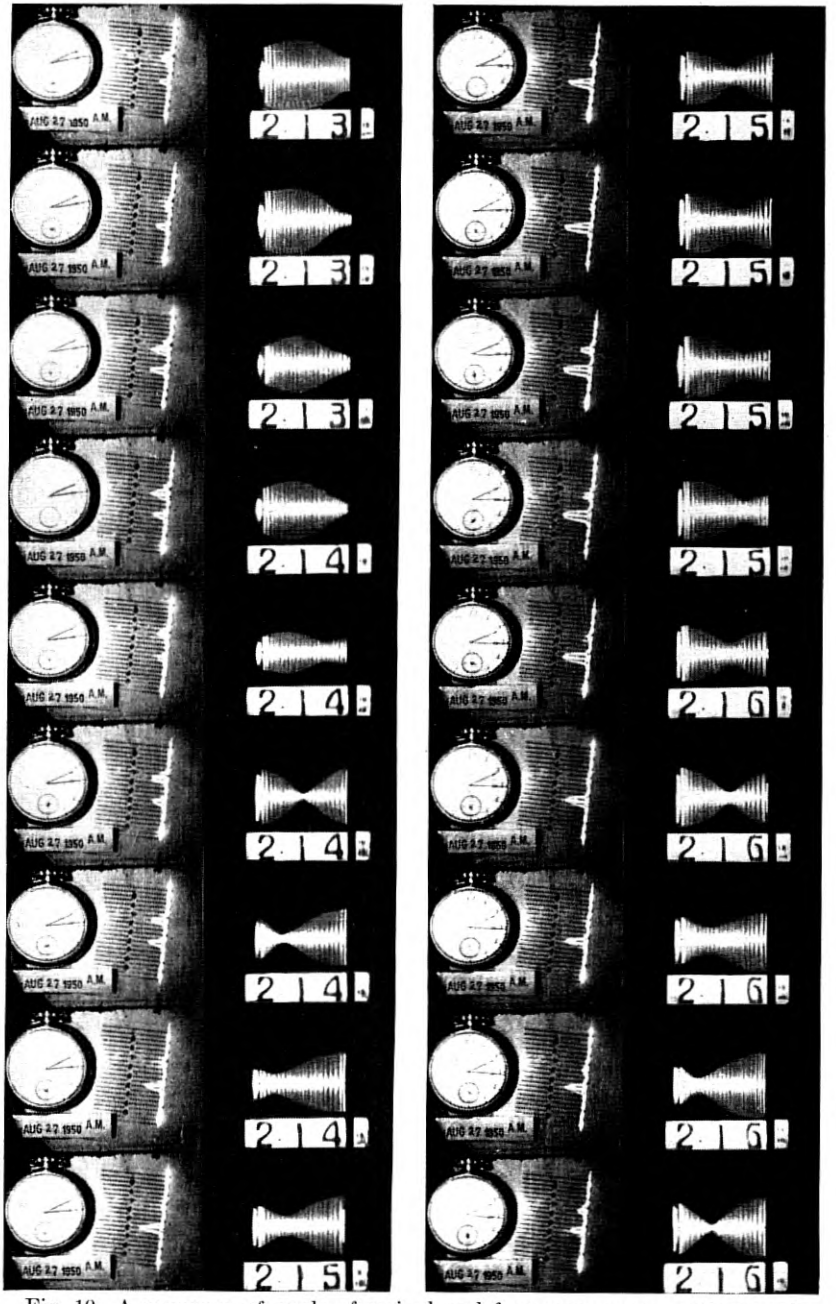
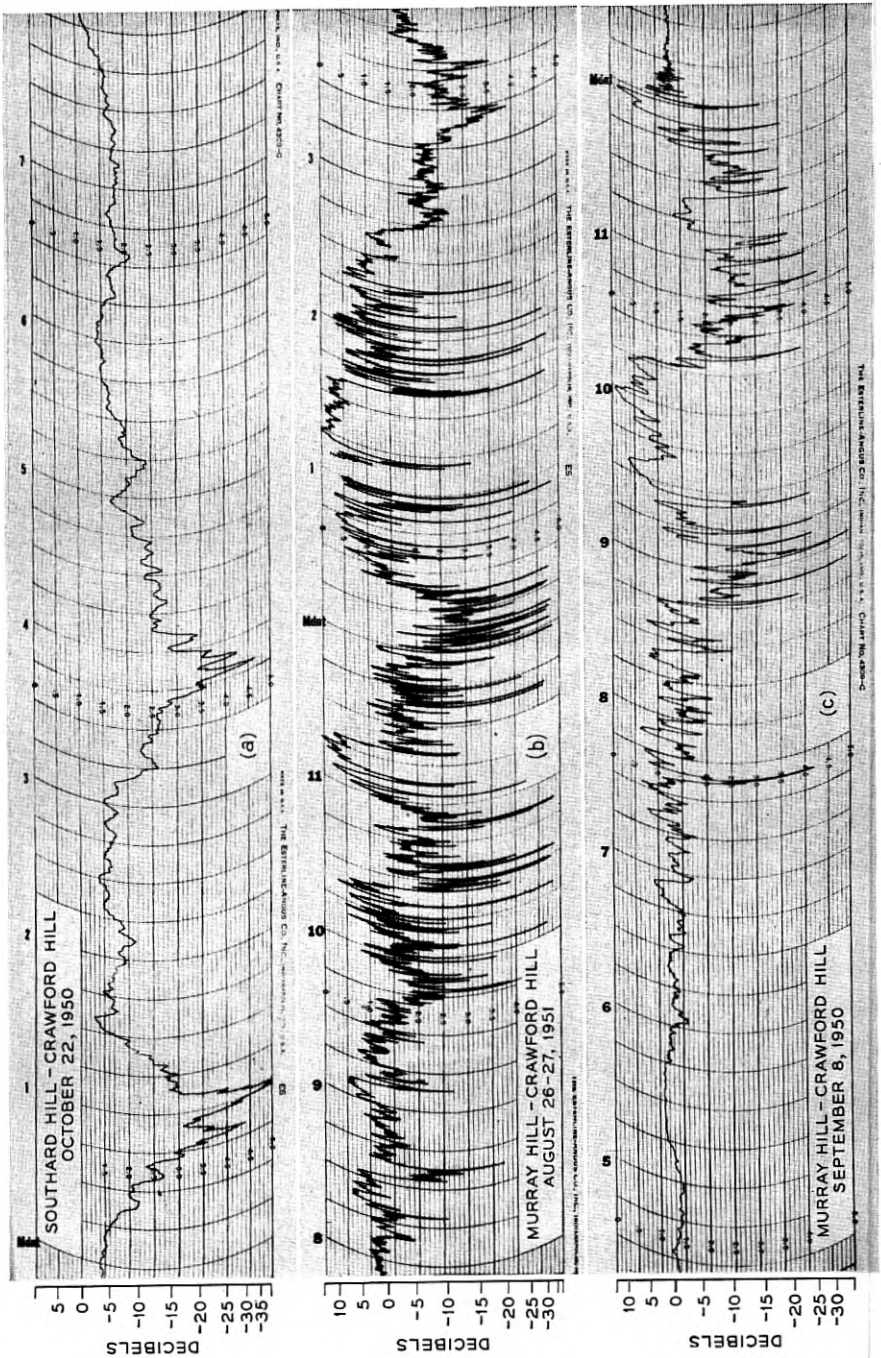


Fig. 10—A sequence of angle-of-arrival and frequency-sweep patterns taken at ten-second intervals.


SOUTHARD HILL – CRAWFORD HILL
OCTOBER 22, 1950
(a)
MURRAY HILL – CRAWFORD HILL
AUGUST 26-27, 1951
(b)
MURRAY HILL – CRAWFORD HILL
SEPTEMBER 8, 1950
(c)
DECIBELS
DECIBELS
DECIBELS

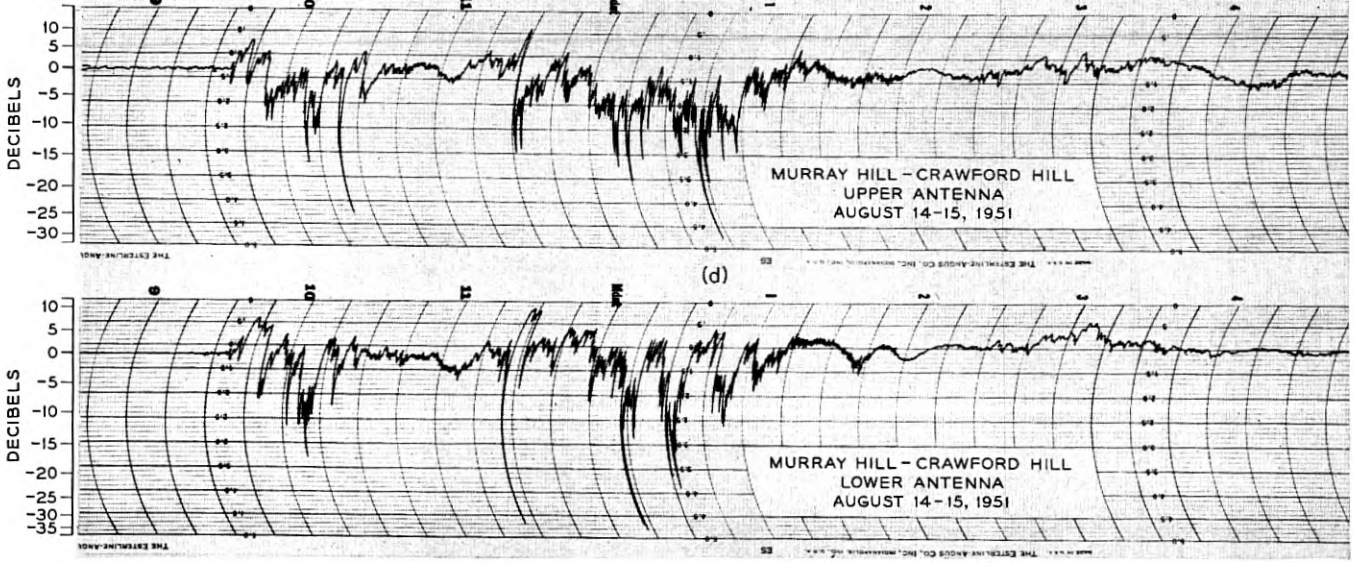
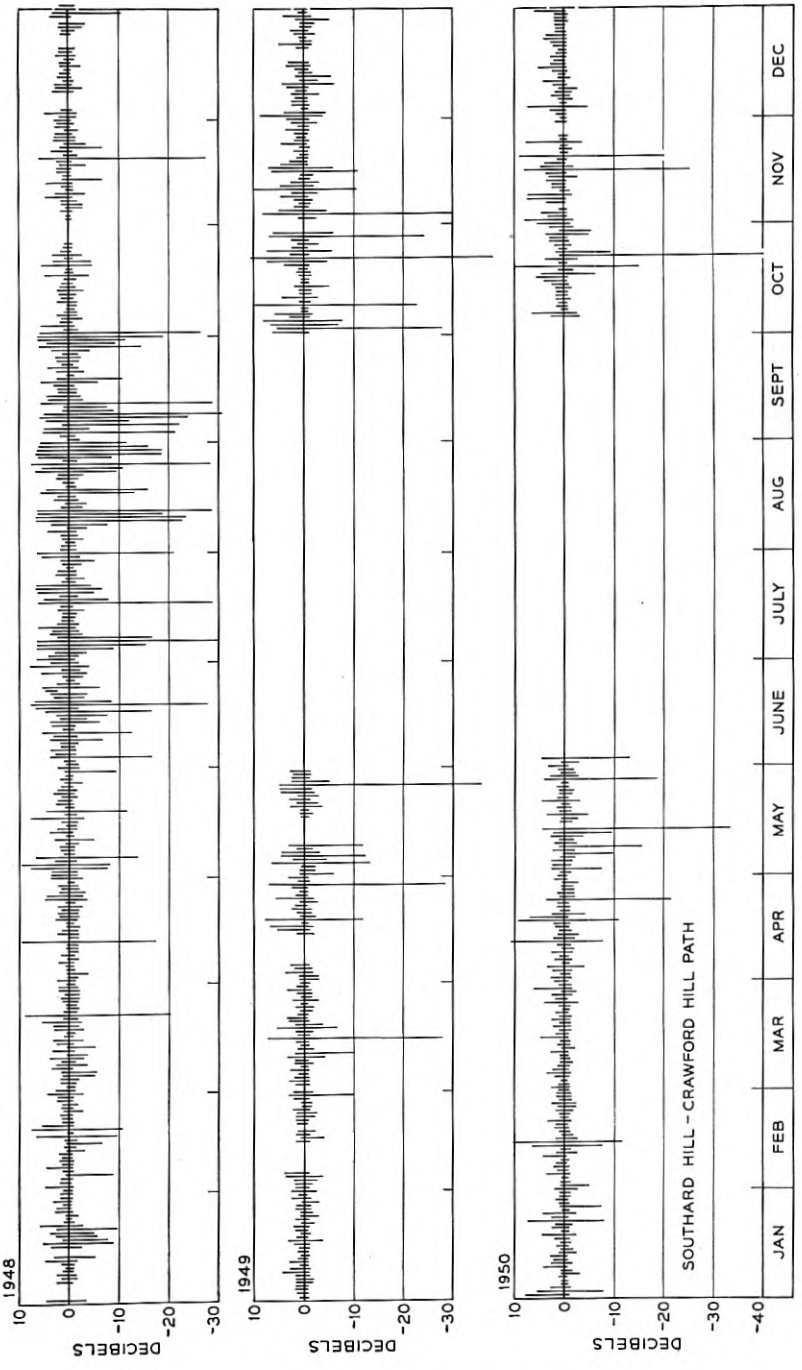


Fig 11.—Typical recordings of 4195-mc transmission during severe fading conditions. The zero decibel line represents the normal daytime (free space) level. (a) Substandard conditions on the Southard Hill-Crawford Hill path. Night of Oct. 21–22, 1950. (b) Multiple path transmission on the Murray Hill-Crawford Hill path. Night of Aug. 26–27, 1950. (c) Multiple path transmission on a night when abnormal water reflections were present on the Murray Hill-Crawford Hill path. Night of Sept. 8–9, 1950. (d) Simultaneous recording of the outputs of two similar antennas; vertical spacing of 30 feet. Night of Aug. 14–15, 1951.


1948
1949
1950
10
0
-10
-20
-30
-40
DECIBELS
SOUTHARD HILL - CRAWFORD HILL PATH
JAN
FEB
MAR
APR
MAY
JUNE
JULY
AUG
SEPT
OCT
NOV
DEC

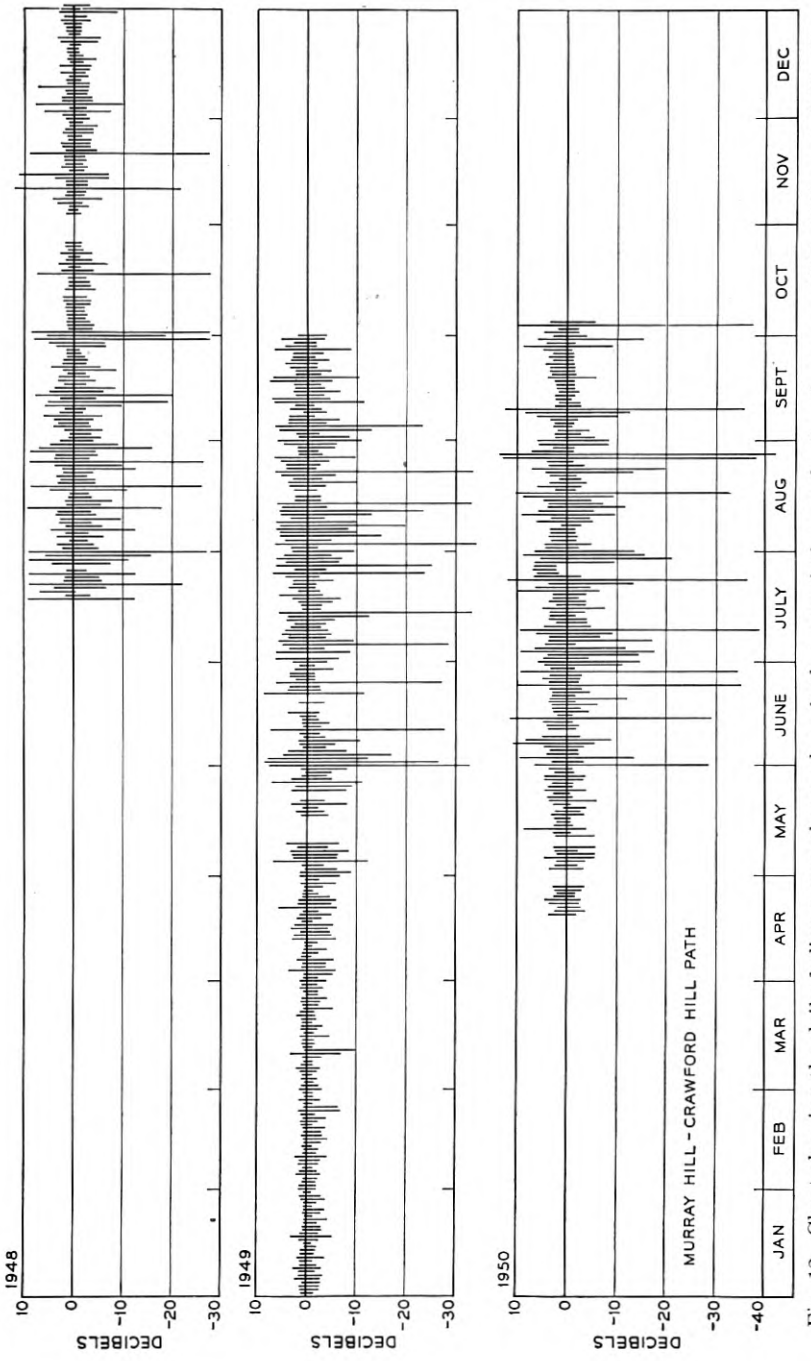


Fig. 12—Chart showing the daily fading range observed on both transmission paths for the years 1948, 1949, and 1950. The vertical lines terminate on points indicating the highest and lowest signals observed during the twenty-four hour period of noon to noon.

of short duration sometimes observed when the average signal level was depressed by the mechanisms of Figs. 3(b) or 3(d) were found to be frequency selective.

Some of the studies described in this paper were made with vertically polarized waves and some with horizontally polarized waves; at times, 45° polarization was used. In so far as it was possible to determine, the propagation characteristics of both paths were independent of the polarization used.

No meteorological soundings were made in connection with this work. Considering the rapid changes usually observed with the angle-of-arrival and frequency sweep apparatus, it is doubtful that meteorological measurements made in the usual manner would show much correlation with the radio observations except, perhaps, in a general way. The sequence of pictures in Fig. 10 is included to show how the angle-of-arrival and

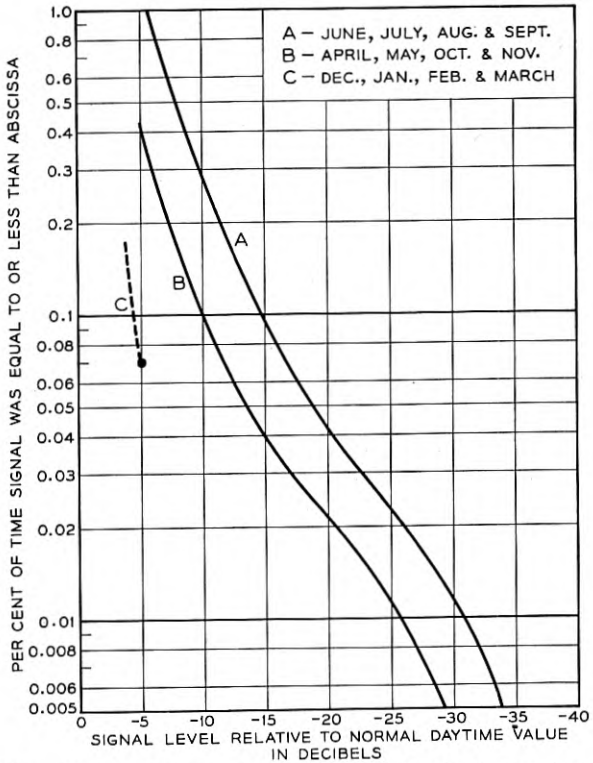


Fig. 13—Time distribution curves of the signal levels observed on the Murray Hill-Crawford Hill path. Data of 1947, 1948, 1949 and 1950.

frequency-sweep patterns change with time. These pictures were taken at 10-second intervals. On this occasion there was good correlation between the angle-of-arrival and frequency-sweep data. Such was not always the case, however, and considering the wide difference in operating frequencies, 24,000 mc and 4000 mc, instantaneous correlation should not necessarily be expected.

Although all the studies described in this paper were made on the two local paths, the results are compatible with propagation measurements made by another group in the Laboratories during a survey for the transcontinental radio relay system.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the contributions of W. M. Sharpless who, for some time, was associated with this work; also to acknowledge

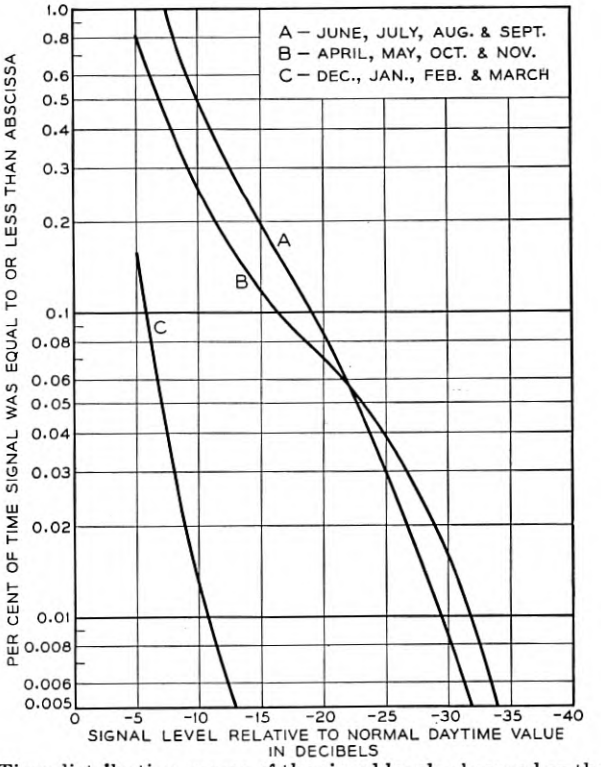


Fig. 14—Time distribution curves of the signal levels observed on the Southard Hill-Crawford Hill path. Data of 1947, 1948, 1949 and 1950.

the full time assistance of R. A. Desmond and the part time assistance of L. R. Lowry and S. E. Reed. All the work was done under the guidance of Dr. H. T. Friis.

APPENDIX

This appendix is included to illustrate some of the characteristics of the propagation as shown by the recordings of 4195-mc signal levels. Fig. 11 is a reproduction of some typical records obtained during severe fading periods. Fig. 11(a) is an example of transmission during the substandard conditions illustrated by the ray diagram of Fig. 3(d). Fig. 11(b) is typical of multiple-path type fading in which the signal components arrive from elevated angles as shown in Fig. 3(a), while Fig. 11(c) was recorded on a night when, for a time, there were abnormal reflections from the water of Raritan Bay on the Murray Hill path, see Figs. 3(c), 6(f) and 7(c). The records of Fig. 11(d) show how the outputs of two similar antennas, spaced vertically about 30 feet, differ in regard to the deep fades of short duration.

The chart of Fig. 12 shows how the fading varies with the time of year. On this chart, the vertical lines represent the extremes in signal level observed during the twenty four hour period from noon to noon. The large signal variations are concentrated mainly in the summer months.

The time distribution of the signal levels recorded on the Murray Hill-Crawford Hill path are shown in Fig. 13. Each of the curves is for a four-month period: the period of least fading, December, January, February and March; the period of most fading, June, July, August and September; and the in-between period consisting of April, May, October and November. Data obtained in the years 1947, 1948, 1949 and 1950 are included. Fig. 14 shows similar data for the Southard Hill-Crawford Hill path. The hump in the time distribution curve for the months of April, May, October and November is due to substandard conditions, illustrated by the ray diagram of Fig. 3(d) and the typical record of Fig. 11(a), which affected transmission on this path during several nights in October, particularly in the years 1947 and 1950. When it occurred, this type of transmission usually persisted for a period of several hours.

Propagation Studies at Microwave Frequencies by Means of Very Short Pulses

BY O. E. DE LANGE

(Manuscript received March 27, 1951)

Microwave pulses with a duration of about 0.003 microseconds were transmitted over a 22-mile path from Murray Hill, N. J., to Holmdel, N. J., in order to determine the effects of the transmission medium upon such pulses. During "fading" periods multi-path transmission effects with path differences as great as 7 feet were observed, as well as some other effects. A microwave frequency of 4000 megacycles was employed.

INTRODUCTION

This experiment was set up with two main purposes in view: First, as a means of studying microwave propagation, especially with regard to multi-path transmission effects and second, to determine the effect of a transmission path upon the shapes of very short pulses, particularly to learn what restrictions might be imposed upon minimum pulse length or spacing between pulses by distortions produced in the transmission medium.

In regard to multi-path transmission the pulse method seems to be the most straightforward way of studying such effects. For example, if there is transmission by more than one path, and if the pulses are sufficiently short in comparison to the path length differences involved, then there will be received a separate pulse for each path. Under these conditions the number of paths involved, path length differences and other information become directly evident. If pulse duration is too great with respect to the path differences involved, the pulses received via the various paths will overlap in time and the resultant multi-path effect will be pulse distortion rather than reception of individual pulses. This situation is much more difficult to analyze.

TRANSMISSION PATH

The transmission path is the same as that used by A. B. Crawford for microwave propagation studies by means of the frequency sweep method,

i.e., the path from Murray Hill, N. J., to Crawford's Hill (near Holmdel), N. J.[1] The path length is approximately 22 miles, and is partly over water and partly over rough land terrain. The frequency sweep studies had indicated that the path differences involved in multi-path fading were of the order of one or two to about seven feet. In terms of delay times this means differences of about 1 to 7 millimicroseconds. In order to resolve the paths when the path differences were only one or two feet, we should have liked to have pulses of about 1 millimicrosecond duration. Because of the difficulties involved in generating, amplifying

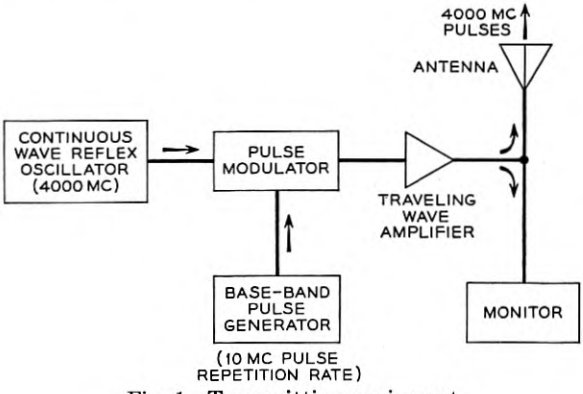


Fig. 1—Transmitting equipment.

and detecting such short pulses, we accepted pulses which, when displayed on our final indicating equipment, had a length of 3 millimicroseconds at half amplitude. (About 6 millimicroseconds at the base.) In free space this pulse would be just about 6 feet long at the base.

TRANSMITTING EQUIPMENT

The transmitter was mounted on top of a 100-foot tower at Murray Hill. As can be seen from Fig. 1, it consisted of a c-w reflex oscillator operating at 4000 megacycles, a baseband pulse generator, a modulator, or gate, for modulating these pulses on the microwave carrier, a single stage traveling-wave amplifier and finally a horn antenna. Approximately one watt of power was obtained from the transmitter at the peaks of the pulses. The antenna area was 25 square feet and its gain 32 db above that of a dipole. A pulse repetition frequency of 10 mc was employed.

[1] A. B. Crawford and W. C. Jakes, Jr., "Selective Fading of Microwaves," *Bell System Tech. J.*, **31**, Jan. 1952, pp. 68–90.

RECEIVING EQUIPMENT

The receiving antenna, a large horn, was mounted between two poles guyed for support. It had an aperture of about 90 square feet and a gain of approximately 38 db over a dipole. The receiver circuit is shown in Fig. 2. About 60 db of gain at 4000 mc was provided by either two or three stages of traveling-wave tube amplifier depending upon the gain of the particular tubes used. It was necessary to provide very good shielding and also careful filtering of all power leads to eliminate the tendency for this amplifier to sing. The amplifier fed two crystal detectors through a hybrid tee junction. Each detector employed a silicon crystal of the IN23B type.

Two indicator circuits are shown in Fig. 2. These circuits are very similar except that one employed a vertical amplifier coupled to a Dumont 5XP2 CRO tube, whereas in the second the baseband output of the crystal was fed directly onto the deflection system of a traveling-wave type of CRO tube. The latter CRO tube, which has been described by J. R. Pierce in the November, 1949, issue of *Electronics*, has a very high deflection sensitivity and is used with a microscope to enlarge its trace; hence, no amplification was required between it and its driving crystal. The deflection system of this tube has a bandwidth of 500 to 1000 mc. The micro-oscilloscope was provided primarily for photographing pulses by means of a 35-mm camera attached to the microscope. (Exposure time was 5 to 15 seconds. The time recorded for each picture corresponds

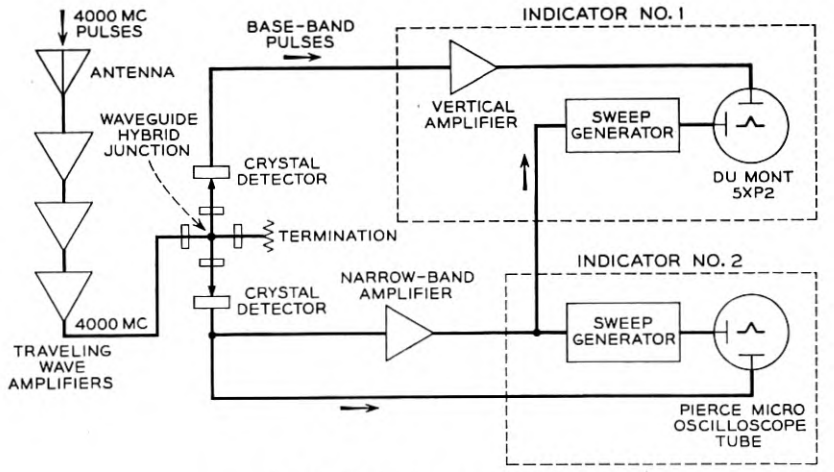


Fig. 2—Receiving equipment.

to that at the end of exposure.) A second microscope made it possible to view and to photograph the screen of the tube simultaneously. The general procedure was to observe continuously during periods of disturbed transmission, taking pictures at regular intervals of 5 to 10 minutes. When conditions were seen to be changing rapidly, pictures were taken much more frequently. The large oscilloscope with its vertical amplifier had a bandwidth of about 150 mc and hence caused some deterioration of the pulse. It, however, was less tiring than the small scope, especially for long periods of observation and was watched to follow the general trend of events. It was capable of resolving the pulses resulting from two-path transmission when the path differences were large.

The sweep circuits for the two indicator oscilloscopes were practically identical. The horizontal sweep voltage for each consisted of the linear portion of a sine wave which was generated by a c-w oscillator operating at one third of the pulse repetition frequency of 10 mc. Each oscillator was synchronized with the incoming pulses by means of a 10-mc voltage derived by amplifying the pulse energy through a narrow band amplifier. This circuit provided very satisfactory synchronization even during the times when signal amplitude was so low as to produce a very poor signal-to-noise ratio. Timing markers were provided on each roll of film by periodically photographing a series of pulses spaced by an interval of 9 millimicroseconds.

RESULTS OF THE EXPERIMENT

The picture at the left of Fig. 3 shows the transmitted pulse. The right-hand picture shows the received pulse under what were considered to be normal transmission conditions. It is seen that, except for the addition of noise and widening of the pulse due to passage through the amplifiers and other equipment, the pulse shape is unaffected. The time calibration on this and the following photographs are in millimicroseconds, each mark representing one millimicrosecond (0.001 μs).

During the summer of 1950, when this experiment was in progress,

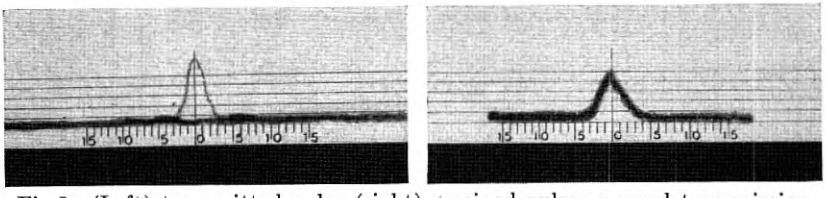


Fig.3—(Left) transmitted pulse (right) received pulse—normal transmission.

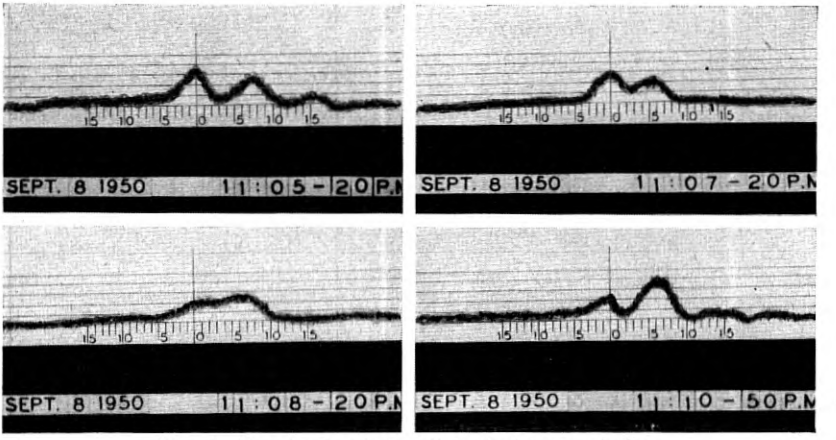


Fig. 4—Received pulses—disturbed transmission.

there was comparatively little fading over the path in question at microwave frequencies. There were, however, a few nights of considerable activity and some interesting results were obtained.

The series of pictures on Fig. 4 show one good example of multi-path transmission where the path length difference was great enough to produce complete resolution of pulses. At 11:05-20 there are two pulses, each 7 millimicroseconds wide at the base and with their peaks just 7 millimicroseconds apart; in other words, the path difference was just sufficient to produce two pulses with no overlap. The pulse at the left is presumably coming by the main path and that at the right from some second path resulting from bending of the rays caused by atmospheric effects. At 11:07-20 the second path appears to have shortened, resulting in a path difference of only about 5 millimicroseconds. This may actually have been due to a change of length of the second path or it may have been due to distortion of the second pulse by energy coming by way of a third path. The pictures taken at 11:08-20 and 11:10-50 show evidence of transmission by a third path. In the first of these, for example, the width of the disturbance at the base line indicates the presence of the two original pulses spaced 7 millimicroseconds apart but the midpoint of the two no longer falls to the base line as was the case in the first picture. This could be accounted for by the presence of a third pulse coming over a path whose length was somewhere between that of the other two. Conditions obtaining at 11:10-50 could also be accounted for by the presence of pulses from three paths, that is, energy coming by way of a third path might cancel part of one pulse and at the same time add to the other. This could account for the fact that the spurious pulse is larger than the

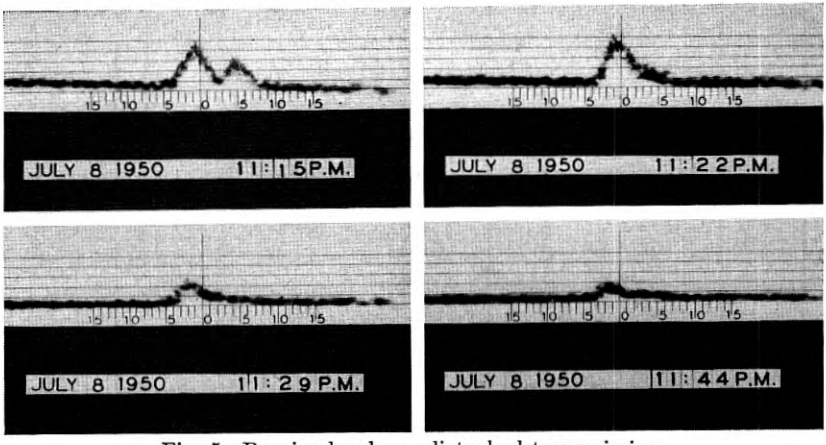


Fig. 5—Received pulses—disturbed transmission.

normal one. It is also possible that more than three paths were involved. On a number of other occasions the pulse coming by way of the second path appeared to be of greater amplitude than the one coming by the main path. This same effect has been observed by Mr. Crawford and his colleagues on the angle of arrival equipment.

Information obtained from the above set of pictures shows that for a time-division multiplex system using the length of pulse used here (7 millimicroseconds at the base) and operating over this path, pulses would have to be spaced a minimum of about 14 millimicroseconds apart if it were desired to avoid distortion at all times. If very much shorter pulses were used the spacing might be reduced to 9 or 10 millimicroseconds. However, the 7-foot path difference indicated by these pictures is about the maximum ever observed and occurs rather infrequently so that if somewhat closer spacings were employed troubles would result only a small percentage of the time.

The next series of pictures, Fig. 5, taken July 8, show an example of a more common type of multi-path transmission. Here the path difference is apparently less than for the last series. At 11:15 there are two distinct pulses with an apparent path difference of about six feet (6 millimicroseconds) if judged from the spacing between the peaks of the pulses. However, from the length of the disturbance at the base line, which we consider a better criterion, the path difference was more nearly four feet At 11:22 distortion of the trailing edge of the pulse was the only indication of a second path. For the pictures taken at 11:29 and 11:44 the path difference is sufficiently small that there is almost complete cancellation of pulses, only the leading portion of each pulse being present.

On the 11:44 picture there is just a trace of a second pulse. The next set of pictures (Fig. 6) were taken a little over an hour later on the same night and show about the same conditions, that is, pulse amplitude and shape change and other evidence of the presence of a second pulse delayed about 2 to 3 millimicroseconds.

On the night of October 2, fading, which was apparently due to transmission by way of two paths with little path difference, was observed. Some of the results are shown on Fig. 7. At 7:49 two distinct pulses are evident, there being 6 millimicroseconds between their peaks. One might conclude from this that there was a second path about 6 feet longer than the main path but the total length of the disturbance along the base line and the shapes of the pulses indicate that the actual path difference was about 2 to 3 feet.

Apparently we had here two pulses of r-f energy overlapping in time and involving a large number of frequencies. These pulses are capable of interfering with each other in a rather complicated manner, it being possible for some frequencies to add and others to cancel at the same time, depending upon their relative phases. Phase relationships of course depend upon frequency and path length differences. As a result pulses may be distorted and have their peaks shifted about by a considerable amount. We must, therefore, realize that the first picture of Fig. 7 does not really represent two distinct pulses as appears to be the case, but actually shows the resultant interference pattern of two overlapping pulses. Since a change of path difference of only about one and one-half inches is enough to produce a 180° change in relative phase at 4000 mc, it is not

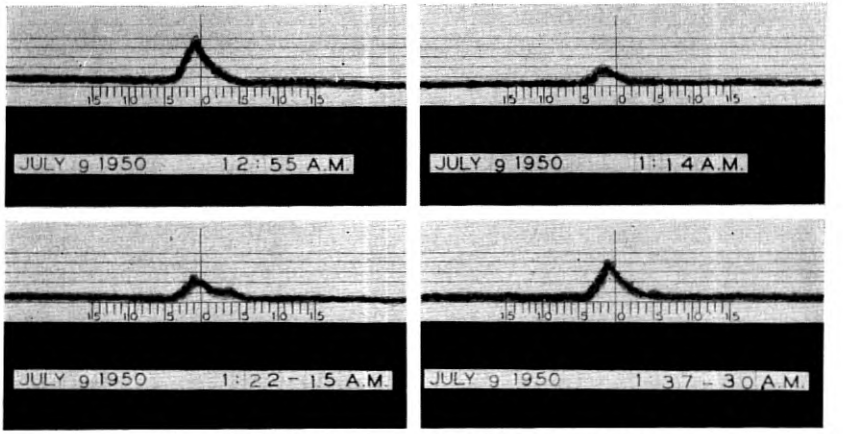


Fig. 6—Received pulses—disturbed transmission.

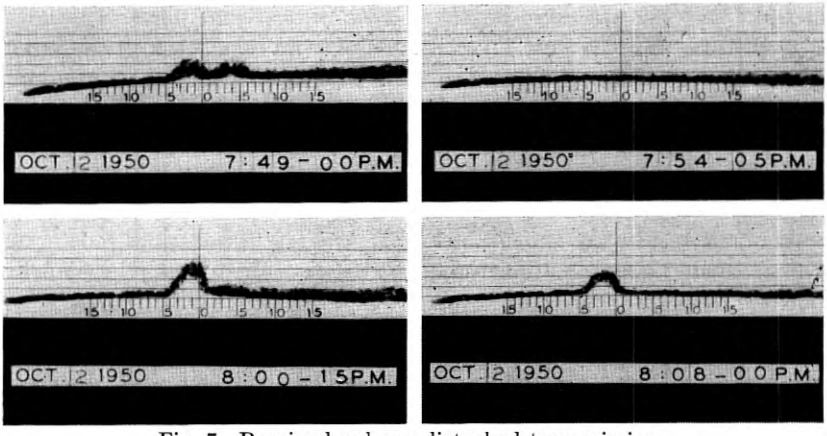


Fig. 7—Received pulses—disturbed transmission.

at all surprising that pulse shapes and amplitudes change very rapidly at times.

Looking again at the photograph, Fig. 7, we see that at 7:54-05 there was a complete fade as far as our system is concerned. To produce this degree of cancellation the path difference must have been very small though still sufficient to give a relative phase angle of 180° at the radio frequencies involved. At 8:00-15 and 8:08-00 pulse distortion is the most noticeable effect of the "fading," the pulses being considerably shorter than their normal value. Pictures, not shown here, taken between 7:54 and 8:08 show definite evidence of two-path transmission with a path difference of 2 to 3 feet; therefore the pulses of 8:00-15 and 8:08-00 are probably also the result of two-path effects.

The first two pictures of Fig. 8 show another form of pulse distortion observed on a number of occasions. Here the pulse is flattened out on top probably due to energy coming in over a second path differing in length by only one or two feet from the main path. Each time this type of pulse was observed a check was made to be sure that the flattening was not due to overload in our equipment. The pictures presented up to now have all shown comparatively slow changes of conditions. Very rapid changes were, however, quite common. In many cases pulse shape or amplitude changed considerably during the 5 to 15 second exposure time ordinarily used. The picture taken at 2:20-45 A.M. on August 27 is one example of such a rapid change, there being two definite sets of conditions shown on the one photograph. The remaining picture on Fig. 8 shows the pulses used for obtaining time calibration of the system. These pulses were spaced 9 millimicroseconds apart and by adjust-

ing sweep expansion so that succeeding pulses fell on proper parts of the scale and by keeping this expansion constant, it was possible to obtain a calibration.

TWO-PATH SIMULATOR

As an aid to interpreting the results obtained from the above experiment, particularly when the two pulses overlap and interfere, a circuit was set up in the laboratory to simulate two-path transmission. The equipment, as shown on Fig. 9, consisted of a wave guide hybrid junction with the r-f pulse energy being fed into the E plane arm. To each side arm was connected a variable attenuator in series with a few feet of wave guide fitted with a short circuiting plunger. Waves reflected from these two plungers recombine in the H plane arm where the detector is located. There are two separate paths through the hybrid as follows: (1) Input, side arm A, reflecting plunger A, side arm A to output. (2) Input, side arm B, reflecting plunger B, side arm B to output. By adjusting the attenuator in either branch the amplitude of the signal transmitted by way of that branch could be adjusted. In the same way by adjusting the position of the reflecting plunger in either branch the distance traveled by a signal in traversing that branch could be varied.

If the path lengths were made the same and the amplitudes adjusted to be equal there would be perfect cancellation due to a phase turn-over in the hybrid junction and hence no output from the detector. If one plunger were now left fixed in the above position and the other moved by a quarter wavelength (to produce a total shift of half wavelength or 180°)

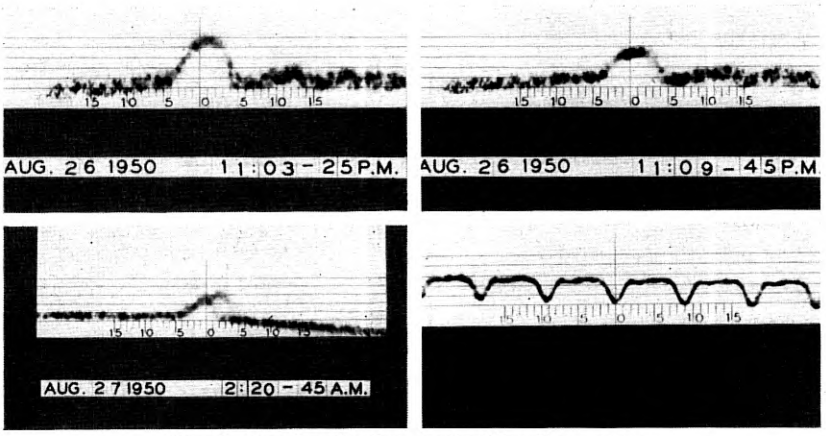


Fig. 8—Received pulses and calibrating pulses.

the output signal would be at maximum amplitude due to addition of energy coming from the two paths. This amplitude is, of course, twice that of the signal from one branch alone. In the experiment the plunger in one branch was left fixed and the attenuator in that branch left set at zero. The path through this branch then represented the normal transmission path for an actual system. The path through the other branch could be made to correspond to spurious paths having different amounts of delay and attenuation simply by adjusting the position of the reflecting plunger and the setting of the attenuator. A series of photographs were taken of pulses resulting from these different amounts of delay and attenuation.

The first three pictures of Fig. 10 were taken with the path lengths exactly equal. When the amplitudes were also equal there was complete cancellation. As the signal in one branch was attenuated the amplitude of the resultant pulse increased until it became equal to that of the original pulse as shown in the third picture. Increasing one path by one-half wavelength brought the signals from the two branches into phase and they added up to double amplitude as seen in the fourth picture. It should be pointed out that although in our experiment we changed delay by 0.36 millimicroseconds in going from the first minimum to the first maximum, in free space a change of delay of only 0.125 millimicroseconds

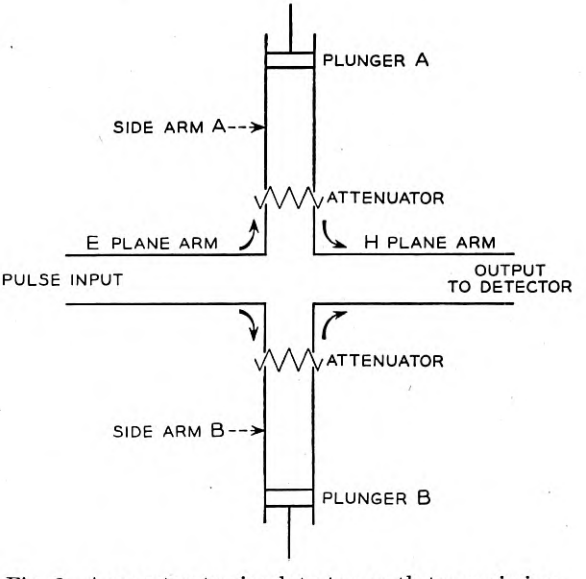


Fig. 9—Apparatus to simulate two-path transmission.

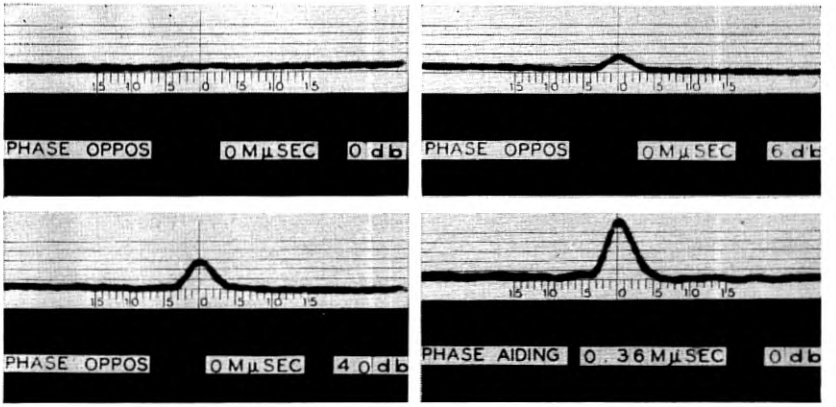


Fig. 10—Simulated two-path transmission.

would be required. The discrepancy lies in the large ratio between the phase velocity and group velocity in the wave guide used whereas in free space this ratio is, of course, equal to unity. In free space the amount of delay required to go from a maximum to a minimum signal corresponds to a change of path difference of only about one and one-half inches. With only this slight shift required to change conditions from those shown by the first picture of Fig. 10 to those shown by the last, it is not at all surprising that the received signal is very unstable during time of multi-path transmission.

Fig. 11 shows the effect of changing relative phases in 90° steps while keeping the amplitudes equal to each other. It is seen that even with the carriers in direct opposition cancellation is far from complete due to the

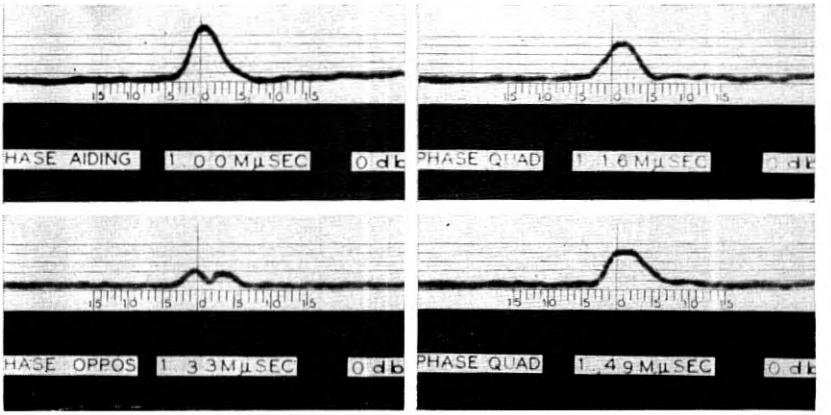


Fig. 11—Simulated two-path transmission.

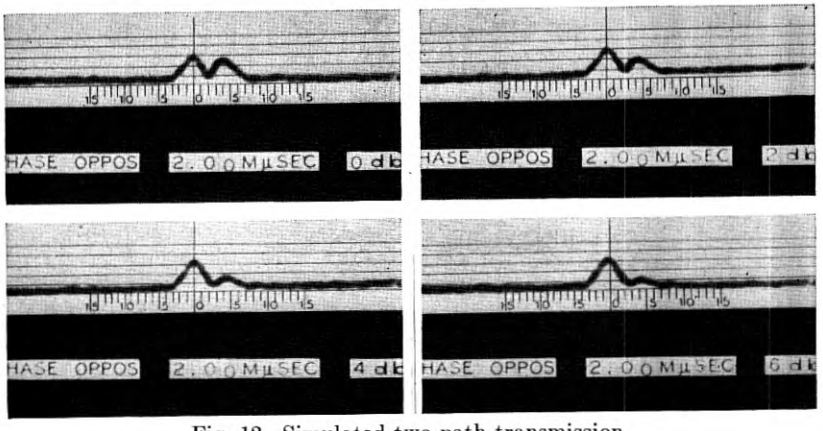


Fig. 12—Simulated two-path transmission.

relative delay between the two component pulses. Flat topped pulses seem to be characteristic of conditions where the two carriers are in phase quadrature and about equal in amplitude.

Fig. 12 shows a set of conditions with a constant delay difference of 2 millimicroseconds (corresponding to a path difference in free space of about 2 feet). For the pulses shown on Fig. 13 there was a constant delay difference of 7.34 millimicroseconds, enough to provide complete separation of the pulses. The carriers were in phase opposition but with this amount of separation there is no overlap of pulses and the results would have been the same if the phase had been aiding. Any increase of delay

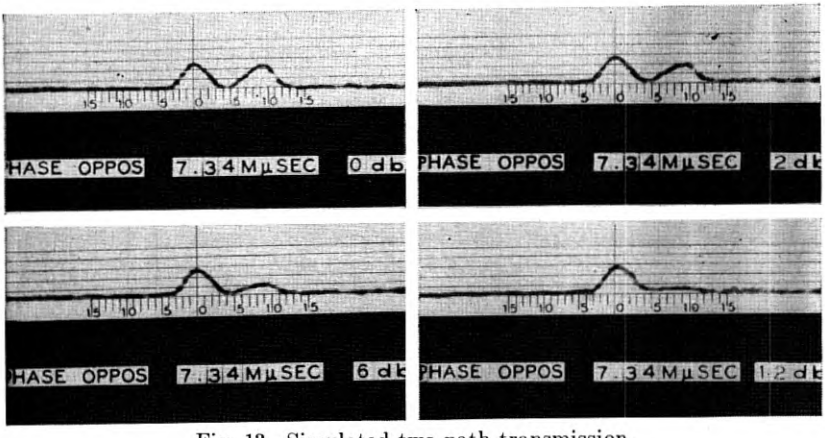


Fig. 13—Simulated two-path transmission.

beyond this point results only in moving the pulses farther apart and has no effect upon pulse shape or amplitude.

The experimental set up just described proved to be somewhat unsatisfactory since it was not possible with it to produce phase opposition between the two carriers without having zero delay difference between the two paths or a difference of at least 0.66 millimicroseconds. For the length of pulse used this latter amount of delay difference is sufficient to prevent anything like complete cancellation of pulses. In fact the amplitudes of the two resultant peaks are only about 12 db below the peak amplitude of the original pulse. From this we know that for the natural path any fade which appeared to be complete must have resulted from path differences of less than 0.66 feet, in fact from differences of less than about one-half foot.

SUMMARY

The pulse experiment results indicate that over one particular path at least there is, at times, transmission of microwaves by at least two, and probably more than two, paths. Path differences involved are from a fraction of a foot up to about seven feet, differences of less than about three feet being the most common. These results agree with those obtained by other methods. These multi-path effects result in bad distortion of very short pulses and even in the presence of entirely separate spurious pulses. These effects put a definite lower limit on pulselength and spacing between pulses in a pulse transmission system. The limit depends upon the amount of distortion which can be tolerated and also upon the percentage of time such distortion can be accepted. No statistical data were recorded.

With the laboratory equipment for simulating transmission over two paths, many of the waveforms obtained over the natural path could be duplicated. There were times, however, when the waveforms received by way of the natural path were too complicated to be explained by transmission by as few at two paths.

ACKNOWLEDGEMENTS

Space does not allow giving credit to all of the many people who contributed to the success of this project. A. F. Dietrich made the mechanical layouts for most of the equipment and supervised its construction as well as assisting in taking data and in many other ways. I also wish to thank J. C. Schelleng and W. M. Goodall for guidance and suggestions. G. M. Eberhardt furnished the traveling-wave-amplifier circuits.

Properties of Ionic Bombarded Silicon

BY RUSSELL S. OHL

(Manuscript received August 23, 1951)

This paper deals with a new and very interesting technique by which the properties of silicon surfaces are altered very materially by bombardment with ions of such gases as hydrogen, helium, nitrogen and argon. The change in rectifying properties has been of special interest but there have been considered also changes in the structural features of the material itself. The effects of bombardment on the rectifying properties are illustrated by a series of characteristic curves systematically arranged to bring out the effects of the several variables of experiment such, for example, as ion velocity, intensity of bombarding current, length of time of bombardment, kind of gas, and the temperature of the specimen during bombardment. The effect of bombardment on materials contaminated with impurities is also illustrated. It is of particular practical importance that silicon contaminated with boron to the point where it shows relatively little rectification can be modified by bombardment to make it even better than most unbombarded materials.

Some years ago, the writer discovered that the electrical properties of silicon surfaces could be greatly modified by bombardment with positive ions. The ions in question were generated in a low pressure discharge in some gas, like hydrogen, helium or nitrogen, and after passing through a perforated cathode were accelerated to a suitable velocity before impinging on the surface to be treated. This scheme may be contrasted with other methods subsequently reported for treating germanium[1] in which high-velocity ions were derived from radioactive sources. Preliminary results of the present research were described in a paper entitled *Silicon Transistors*, by W. J. Pietenpol and the writer, presented at an Electronics Conference held at the University of Michigan, June 22, 1950. Since that time exploration has continued with a view both to learning about basic principles and about possible practical applications.

Editorial Note—Since the resurgence of interest in point-contact rectifiers, considerable research has been carried on into the characteristics of silicon and germanium. The author of this paper was a pioneer in this new field of study, as evidenced, for example, by Patent No. 2,378,944, applied for on July 26, 1939, and Patent No. 2,402,839, applied for on March 27, 1941. More recent work has been described in a large number of text books and technical papers such as *Electrons and Holes in Semi-Conductors* by William Shockley, D. Van Nostrand, 1950, and numerous papers by Lark-Horowitz published mostly in Physical Review. The work described in the accompanying paper is a continuation of this long research.

[1] Brattain and Pearson, *Phys. Rev.*, **80**, Dec. 1950.

The present paper gives the results of some more recent experiments made with improved equipment. Also described briefly are some related experiments in which silicon is bombarded with alpha particles derived from radioactive polonium. The overall results of this work indicate rather clearly that with suitable variations of bombarding voltage, target temperature and time of exposure as well as impurity content in the base material, it is possible to prepare to specification silicon surfaces having a wide range of properties. From the materials so treated it has been possible to construct improved forms of signal rectifiers, harmonic generators, transistors, modulators, gating devices and also photo-electric cells. It is particularly significant that the voltage range over which these newer devices can be operated has been greatly extended, thus making them useful in places not previously regarded as possible. Since these new surfaces appear to be readily reproducible in large numbers and since they are electrically tough, chemically stable and show no unsatisfactory temperature or aging effects, it would appear that bombarding techniques should have considerable practical value.

This paper is concerned mainly with the practical aspect of ion bombardment of silicon, namely its effect on the voltage current characteristics at low frequencies. Equally important, perhaps, are its theoretical aspects, particularly with regard to the interpretation of the rather pronounced changes in the properties in light of presently-accepted views of solid-state physics. These aspects are not covered in this paper.

METHOD

The bombardment process referred to above consists of exposing the silicon surface to ions that have previously been accelerated to energies in the range from about 100 electron volts to about 30 kilo-electron-volts. A recent setup is illustrated in Fig. 1. The electrons from a tungsten cathode are accelerated toward a grid which is at a positive potential with respect to the cathode. Many of the electrons pass a short distance beyond the grid and return for ultimate capture. Ionization due mainly to the impacts of electrons with gas molecules takes place in this turn-around region, producing amongst other things positive particles. Electrodes are so proportioned that this ionization is fairly uniform over the grid area.

The silicon specimen to be bombarded is made negative with respect to the filament. This accelerates the positively charged particles toward the target. The latter rests on a graphite plate heated by a coil below, carrying high-frequency currents. A thermocouple with suitable connections to the exterior makes possible an adequate measurement of

temperature. The apparatus will accommodate circular surfaces as large as $1\frac{1}{2}$ inches diameter. The gases from which ions are derived are admitted through the gas inlet. Thus far experiments have been made with hydrogen, helium, nitrogen and argon. The bombarding voltages have as already noted, been varied from 100 to 30,000 volts and the surface temperatures have been varied from about 20°C to 400°C. The effects of these several variables will be discussed more fully below.

SAMPLE PREPARATION

The material to be bombarded is usually prepared in batches of about 300 grams in fused silica crucibles roughly cylindrical in shape.[2] After solidifying, the cast is ground to a cylinder approximately $1\frac{1}{2}$ inches

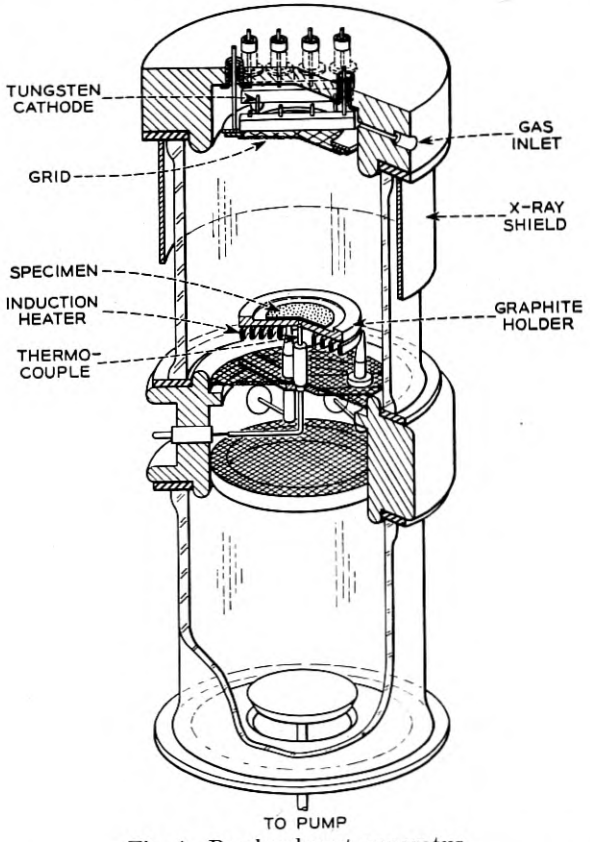


Fig. 1—Bombardment apparatus.

[2] Scaff, Theuerer, and Schumacher, A.I.M.M.E., **185**, pp. 382–392, 1949.

diameter. This has the effect of removing some of the contaminating impurities derived from the crucible as well as providing samples of convenient size. This $1\frac{1}{2}$-inch cylinder is then sliced transversely into thin wafers which subsequently are polished on one side. Except as otherwise noted the material covered by this paper had an impurity content of about 0.1 per cent. The exception will be found in the data of column (a) of Fig. 8.

BOMBARDMENT PROCEDURE

The wafer, as prepared above, is placed in the bombarding apparatus with the rough face contacting the graphite support. The vacuum chamber is sealed by placing the ion generator in position and the whole assembly is evacuated. The sample is then heated to the proper temperature and the desired kind of gas is admitted, the pressure being estimated from the ion current. When stable conditions prevail, the accelerating voltage is applied to the target and the bombardment is carried out for the proper length of time. A convenient current density is 5 microamperes per square centimeter of target area. The target area of our present apparatus, including the silicon and a portion of the graphite support, being 20 square centimeters, the ion current is generally around 100 microamperes. The dosage is sometimes specified in microcoulombs.

After bombardment, the sample is removed from the apparatus and the rough surface is covered with a thin layer of evaporated rhodium. For most of the tests outlined below the $1\frac{1}{2}$-inch diameter wafers were cut into $\frac{1}{8}$-inch squares, a size convenient for testing.

GRAPHICAL REPRESENTATION

In considering the merits of non-linear materials such as silicon, perhaps the simplest and most useful characteristic is the voltage-current relation. If this is plotted to a linear scale, it results in a smooth curve of the general form shown in Fig. 2a. Specific curves obtained in practice may depart widely from that shown but in general, all may be regarded as made up of two semiparabolas, one in the first quadrant and one in the third, joined by a nearly horizontal straight line. For present purposes, we shall further simplify this idealized characteristic by considering it as made up of three straight lines. The first, AB, is associated with the reverse voltage current characteristic. The third, CD, is associated with the forward voltage current characteristic. These two characteristics are joined by the nearly horizontal line, BC. The slopes of these three lines correspond to resistances. The section BC for example, corre-

sponds to a region in which resistance is very high. The points B and C are particularly important for they represent points of inflection where the resistance undergoes rapid change and the material is departing most markedly from Ohm's Law. Ideally they should be sharp but in practice there is usually considerable curvature. Though either inflection point could presumably be used in detection processes, the point to the right of the origin is for practical reasons, usually preferred. Point B defines a voltage E_B at which substantial backward currents flow. It is referred to simply as the *reverse voltage*. In a similar way, point C defines a *forward voltage* E_F. The distance between B and C ($E_B + E_F$) will be referred to as the *inflection interval*. The difference in these quantities ($E_B - E_F$) is also of interest. One-half of this voltage difference is referred to as the *self-biasing voltage*. It is a significant quantity readily measured in practice by noting the d-c voltage across a large condenser placed in series with the crystal and a supply of 60 cycles AC. For detectors, point C should preferably be close to the origin and E_F should

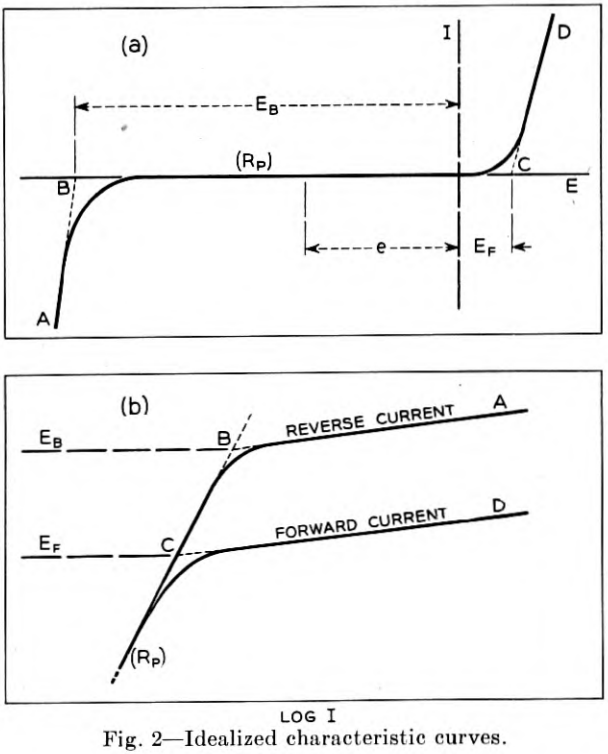


Fig. 2—Idealized characteristic curves.

be small. For certain kinds of voltage limiters, E_F should be large. In either case the inflection interval should be large.

In an alternate graphical representation, see Fig. 2b, voltage-current data are plotted to a log-log scale. This form of representation is of particular value when large ranges of data are to be shown. It is also of value in determining the resistance (R_p) at small voltages. Corresponding points on the two curves shown in Fig. 2 are identifiable by the letters A B, C, and D. Curves of both kinds are used interchangeably to show the effects of the several variables of the experiment.

EFFECT OF CONTACT PRESSURE

In point contact rectifiers,[3] pressure is of considerable importance. Usually the best pressure is a compromise between good electrical characteristics, usually obtainable only with light pressures, and good stability usually obtainable with higher pressures. Experiments have been performed with a range of contact pressures both on bombarded and unbombarded materials. In general, the results are highly variable, particularly in the case of unbombarded material. From this wide range of data, however, two characteristics have been selected that may be regarded as typical for 10-gram and 60-gram pressure. They are shown in Fig. 3 for silicon taken from nearby portions of the same sample. Significant points on these several curves may be compared with their idealized counterparts shown in Fig. 2. Although the samples chosen show somewhat more than the usual intrinsic resistance typical of p-type silicon, the effects of contact pressure are nevertheless regarded as representative. As indicated in Figs. 3a and 3b, the effect of increased contact pressure,[4] particularly in the case of unbombarded material, is of reducing the low voltage resistance, R_p, see Fig. 2b. The more desirable higher resistance is obtainable only with light contact, a condition unfavorable for high mechanical stability. In the case of bombarded material, the effect of contact pressure is less important. Thus it is possible in this case to incorporate in the design higher contact pressures and obtain thereby higher stabilities. For purposes of this paper a contact force of 10 grams has been accepted as standard.

In addition to showing the effect of contact pressure, Fig. 3 shows some overall effects of bombardment. It will be noted, for example, that the effect of bombardment, see Fig. 3b, has been that of shifting the plots of Fig. 3a to the left by several orders of magnitude. Thus the resistance (R_p) is increased by a factor of more than 10,000. It is to be noted also

[3] Scaff and Ohl, *Bell System Tech. J.*, **26**, Jan. 1947.
[4] Realy contact force.

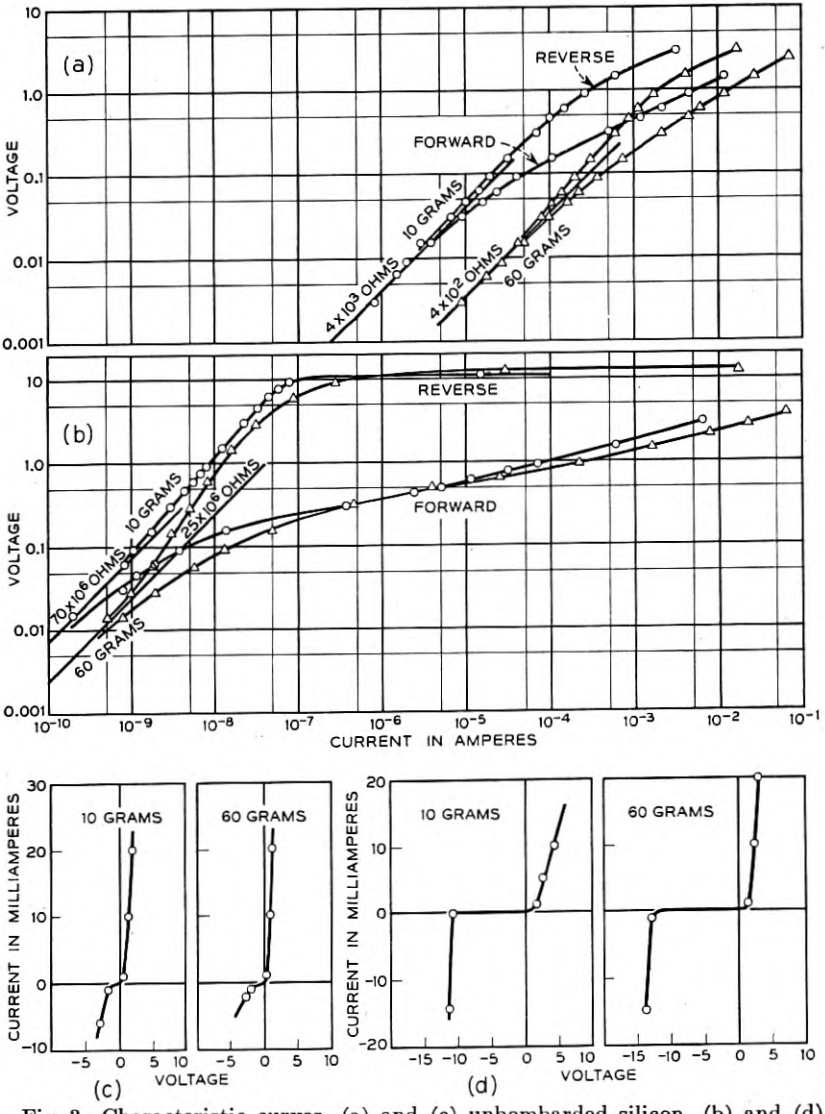


Fig. 3—Characteristic curves. (a) and (c) unbombarded silicon, (b) and (d) silicon bombarded with 30-kv helium ions.

from a comparison of Fig. 3a with Fig. 3b that at the one volt level, the ratio of forward to reverse currents for the unbombarded case is about twenty, whereas that for the bombarded case, is more than 10,000. At other levels the difference is even greater. Referring particularly to Figs. 3c and 3d, it will be seen that one effect of bombardment is that of separating the two significant points of inflection B and C. That is, the inflection interval has been notably increased. This increase is the result of a small increase in the forward voltage and a very substantial increase in the reverse voltage.

EFFECT OF TYPE OF GAS

Four high purity gases were tested as ion sources, namely, hydrogen, helium, nitrogen and argon, having atomic weights respectively of 1, 4,

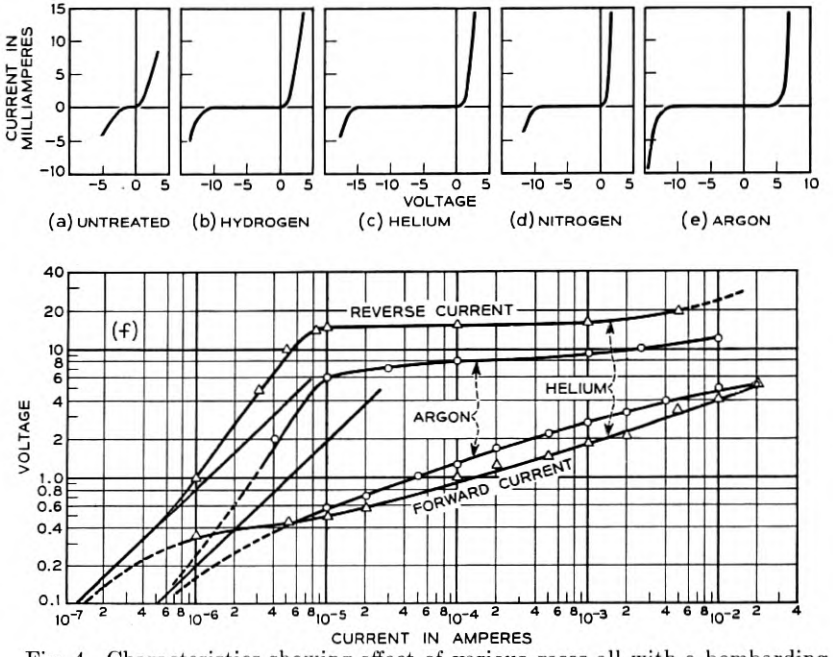


Fig. 4—Characteristics showing effect of various gases all with a bombarding potential of 30 kv.

14 and 40. While all four gases worked very well, helium was the easiest to handle. In the course of the tests, identically prepared samples each $\frac{1}{8}$-inch square, taken from a high-purity silicon melt, were bombarded with ions formed in the particular gas under test. Particular conditions known to be good for producing good rectifier units were adopted as standard for these tests. They corresponded to a total bombarding charge of 600 microcoulombs per sq. cm., a surface temperature of 395°C a contact force of about 10 grams and a bombarding potential of 30 kv.

That the effect of bombardment varies with different gases is seen at a glance from the characteristic current-voltage curves shown in Fig. 4. Figs. 4b to 4e in particular indicate that as compared with an untreated sample, Fig. 4a, the effect of bombardment is in general that already noted of separating the two significant points of inflection, B and C. A rather substantial increase in the forward voltage appears in the case of argon as compared with hydrogen, helium and nitrogen. In contrast with a small increase in the forward voltage resulting from the bombardment of helium, there is a very substantial increase in the backward voltage. Though substantial for all four gases, the effect of bombardment is largest for helium with progressively smaller effects noted respectively for argon, hydrogen and nitrogen. A particular characteristic of helium bombardment, as compared with argon, not readily appreciated from a linear scale, is shown in Fig. 4f. It will be noted that at the one volt level, the ratio of forward to reverse current for helium is about 130 whereas for argon it is about 25. At other levels the difference is even greater. At the moment helium is regarded as a preferred source of ions.

The log-log current-voltage curves show as before that the lowest voltage at which substantial forward currents flow occurs in helium, while the highest forward voltages occur for argon. In a similar way the voltages for substantial reverse currents are highest for helium and lowest for argon. The sharp break in the reverse current characteristic, evident in these cases, has been observed so generally that it is now accepted as typical of bombarded surfaces.

EFFECTS OF TEMPERATURE

Investigations have been made of the properties of silicon surfaces as affected by the temperature at which bombardment was carried out. This has been done not only for surfaces used as rectifiers but surfaces used as transistors and as photo-electric cells as well. In the case of rectifiers, a procedure was adopted similar to that used in the previous tests. Measurements were made at five different temperatures ranging from

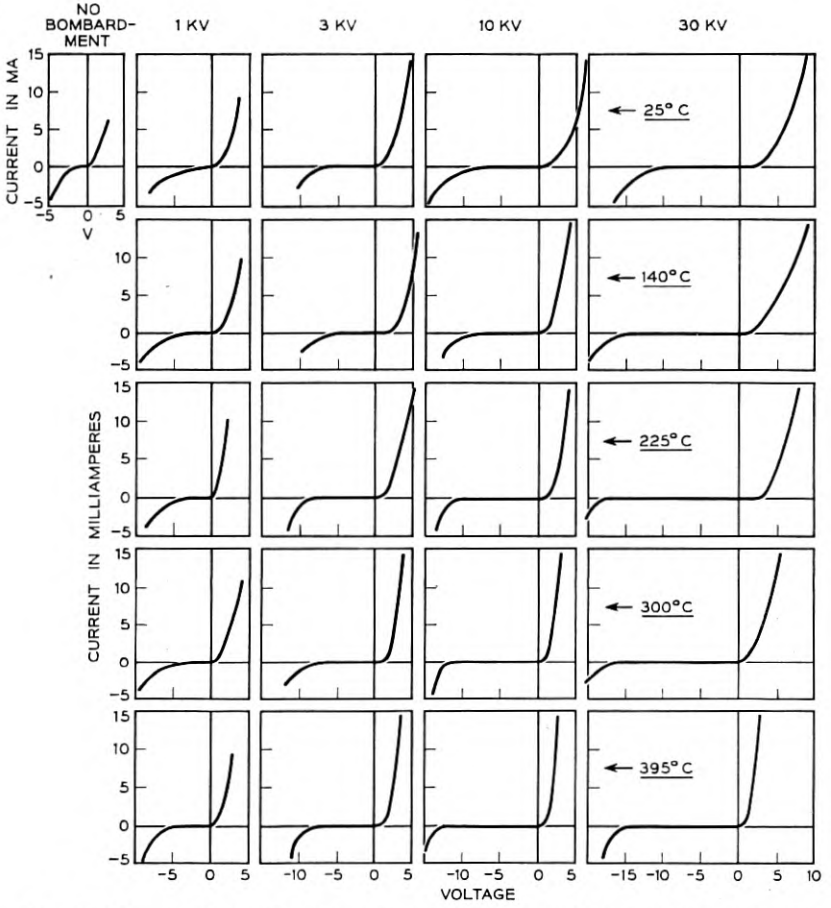


Fig. 5—Characteristics showing effects of voltage and temperature variation.

25°C to 395°C each at accelerating voltages of 1 kv, 3 kv, 10 kv and 30 kv using helium gas. The data so obtained were useful not only for studying the effect of temperature but useful in the studies of the effect of ion velocity as well. The latter will be discussed in the following section.

The results of the above measurements are plotted in Fig. 5. They are further summarized in Fig. 6a. The latter figure, in particular, indicates that as rectifiers, there is little choice of surface temperature between about 250°C and 400°C. It has been found, however, that for temperatures below about 250°C the point contact seems to be more vulnerable to electrical shock.

EFFECT OF ION VELOCITY

The effect of ion velocity (bombarding voltage) has been investigated for several types of silicon. The effects vary with the different types. Typical results are those given in Fig. 5 already referred to. It will be noted from a comparison of the data for a particular temperature, say 300°C, that the principal effect of increased ion velocity is that of increasing the reverse voltage. Values of these reverse voltages E_B and also the

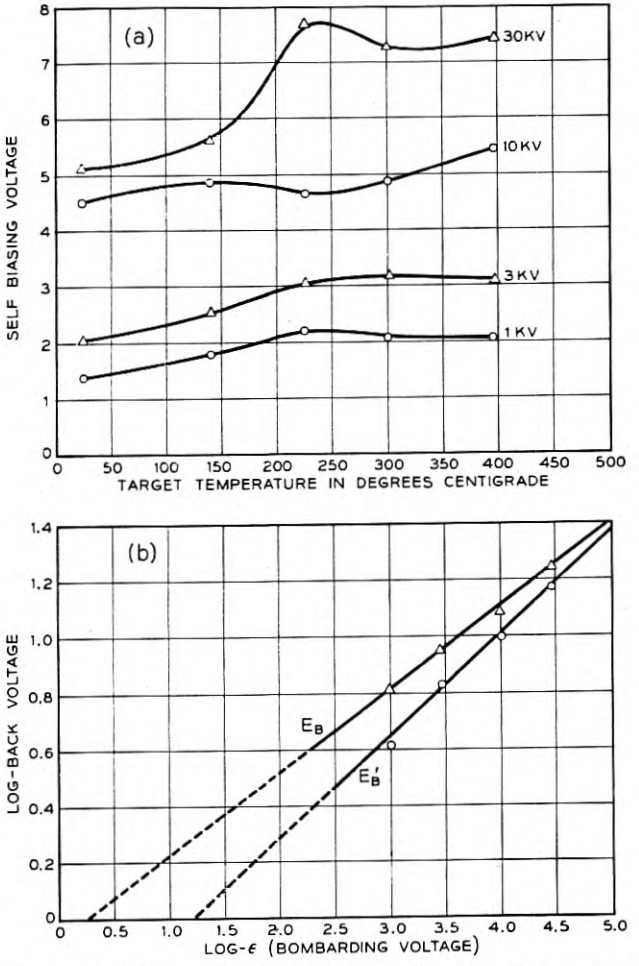


Fig. 6—Summary of data of Fig. 5. (a) effect of temperature and bombarding voltage on self biasing voltage; (b) effect of bombarding voltage on reverse voltage.

TABLE I—EFFECT OF BOMBARDMENT VOLTAGE ON E_B AND E_F

Surface Temp. Deg. C.	30 KV		10 KV		3 KV		1 KV		No Bombardment	
	E_B	E_F	E_B	E_F	E_B	E_F	E_B	E_F	E_B	E_F
395	16.7	1.3	14.0	1.5	10.0	1.5	7.1	1.0		
300	16.5	1.5	12.5	1.2	10.0	2.0	6.5	1.0		
225	18.5	3.6	12.5	1.5	9.6	1.5	5.5	0.3		
140	17.5	2.5	9.8	1.5	6.7	2.5	7.0	1.2		
Mean	17.3	2.0	12.1	1.4	9.1	1.9	6.5	0.9	2.4	0.5
Mean E'_B	14.9		9.7		6.7		4.1			

corresponding forward voltages E_F have been scaled from the above drawings and have been tabulated below. Since they vary only slightly with surface temperature, only their mean values are regarded as significant. Mean values of E_B are plotted in Fig. 6b.

In order to isolate further the effects of bombardment we have subtracted from the mean values of E_B value of E_B for untreated silicon. Thus the curve marked E'_B represents the improvement in backward voltage that has accrued from bombardment alone. This is also tabulated as the mean E'_B in Table I.

EFFECT OF TOTAL CHARGE

Tests have been made to determine the effect of time of bombardment on the rectifying properties of silicon surfaces. In these tests, specimens taken from neighboring regions of the same melt were exposed for progressively longer periods all at the same bombarding potential of 30 kv and the same rate and density of application, 5 microamperes per square centimeter. Representative current-voltage characteristics are plotted in Figs. 7a and 7b for two neighboring regions. The results are summarized in Fig. 7c. The latter show a rather rapid improvement of back voltage E_B with total charge up to perhaps 50 microcoulombs per square centimeter.[5] Thereafter the improvement is small. For purposes of comparison, there is plotted as a vertical line a value of bombarding charge that would account theoretically for one positive ion in each unit crystallographic cell on the surface layer. This suggests that when all surface cells have been penetrated by a single ion, no marked increase in back voltage can be effected.

[5] Specifying results in terms of microcoulombs implies that bombarding effects are independent of the rate of application. This is known to be true only between factor limits of $\frac{1}{2}$ and 2 of the bombarding current.

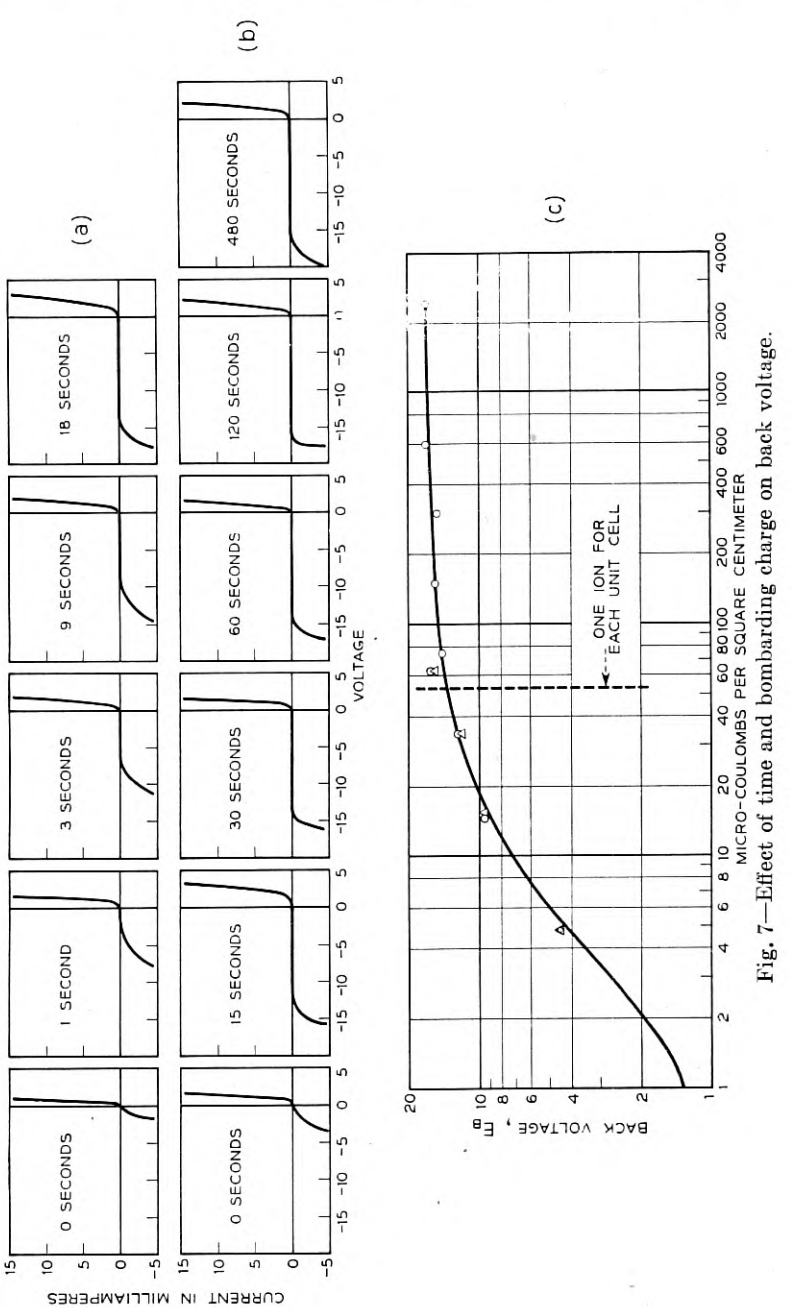


Fig. 7—Effect of time and bombarding charge on back voltage.

EFFECT OF MATERIAL COMPOSITION

Thus far discussions have centered around a single type of high purity material that was regarded as representative. It is of interest to examine the effect on other materials particularly those in which impurities have been added. For this purpose tests were made on comparable samples from four sources all bombarded for two minutes with 5 microamperes of current and each with five representative bombarding voltages. The results are illustrated by the curves shown in Fig. 8. The four columns correspond to progressively higher percentages of impurities beginning with (a) on the left as a material having an impurity content believed to be less than 0.01 per cent. The impurity content of (b) is not known accurately except that it lies between (a) and (c). The material represented by column (c) was produced by adding 0.02 per cent boron[6] to a material illustrated in column (a). The last column (d) was produced by adding 0.1 per cent boron to the material illustrated in column (b).

It is to be noted from Fig. 8 that marked changes in the voltage-current characteristic may be effected by bombardment for all degrees of the impurity content shown. It is especially interesting that in columns (c) and (d) corresponding to materials contaminated with boron to the point where nonlinearity is almost absent, rectification can not only be restored but indeed the product may be made better than the best unbombarded material.

EFFECTS OF ALPHA-PARTICLE BOMBARDMENT

The close relationship between helium ions such as generated above and alpha particles such as emanate from radioactive materials suggests that the latter may be used for the bombardment of silicon surfaces. A few experiments of this kind have been made with results that are not only interesting but possibly useful. For these tests, four sources of alpha particles were obtained. They consisted of $\frac{5}{16}$ inch square pieces of nickel on which had been plated a thin coating of polonium followed by a covering of gold. The initial strength was 4 millicuries per square centimeter. The half-life of polonium is 140 days.

The process of bombardment consisted simply of placing the polished surface of a standard silicon square against the layer of gold and examining the same periodically. Tests of four samples were carried out simultaneously. The results are given in Table II. The data for Sample No. 1 departs so markedly from the mean that it may be disregarded. Since

[6] Boron is a particularly active agent in effecting changes in the properties of silicon.

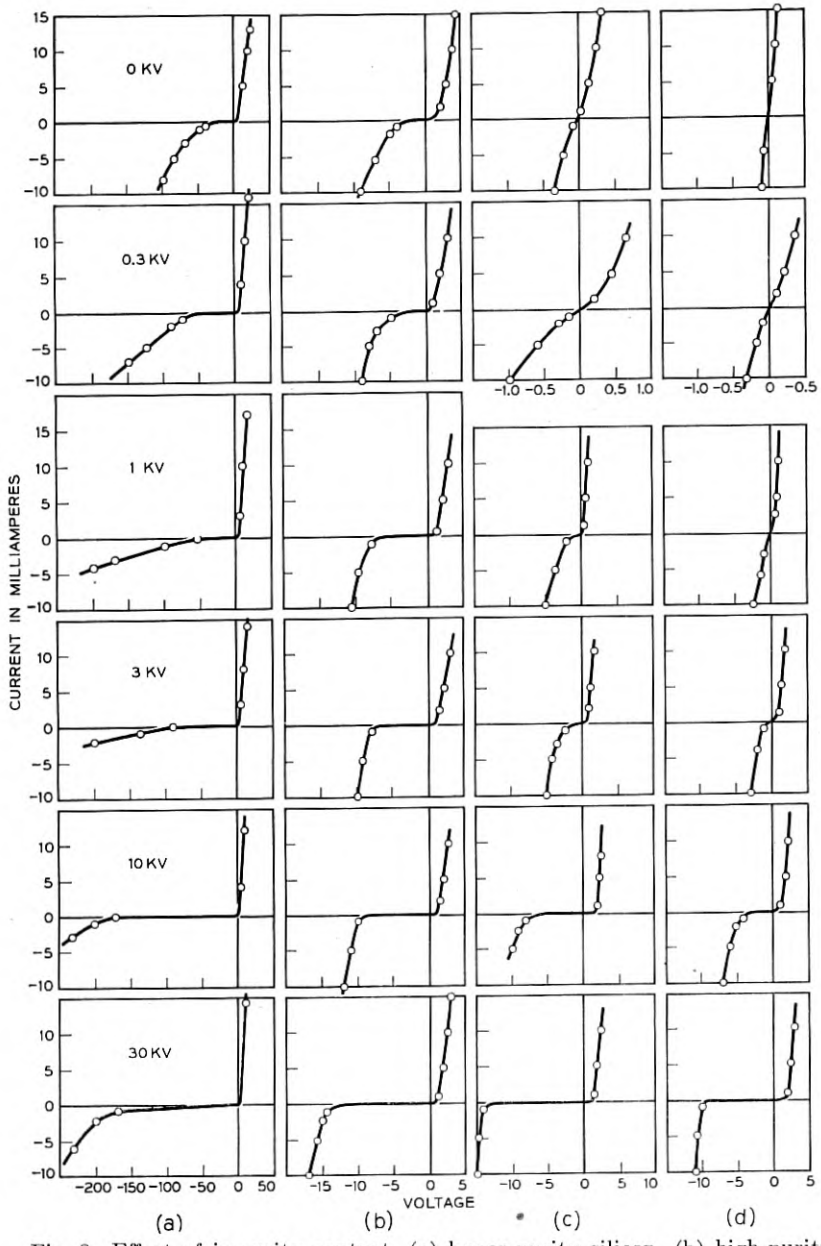


Fig. 8—Effect of impurity content. (a) hyper-purity silicon, (b) high-purity silicon, (c) hyper-purity silicon plus 0.02 per cent boron, (d) high-purity silicon plus 0.1 per cent boron.

the data for self-bias are the result of a direct measurement while those for forward and reverse voltage are transcribed from a cathode-ray plot, they are perhaps the most significant. Samples 1 to 3 represent specimens of increasing degrees of purity.

These alpha-particle bombardment experiments indicate rather definitely that results may be obtained similar to those obtained from bombardment with gaseous ions and, like the ion bombardment, they tend to produce high resistance surfaces.

TABLE II

Days of Exposure		Sample Number			
		1	2	3a	3b
4	Self Bias-Volts	<1	6	180	60
4	Reverse Voltage	<1	20	380	130
4	Forward Voltage	<0.5	<0.5	<1	<1
13	Self Bias-Volts	<1	160	200	190
13	Reverse Voltage	1	360	440	420
13	Forward Voltage	<0.5	<0.5	<1	<1
39	Self Bias-Volts	<1	60	130	100
39	Reverse Voltage	0.5	480	500	500
39	Forward Voltage	<0.2	350	180	180
56	Self Bias-Volts	<1	70	150	110
56	Reverse Voltage	0.5	440	560	560
56	Forward Voltage	0.3	320	320	240
66	Self Bias-Volts			100	85
66	Reverse Voltage			560	520
66	Forward Voltage			200	320

MECHANICAL EFFECTS OF BOMBARDMENT

The marked changes in the electrical properties of silicon imposed by bombardment strongly suggest that bombardment may also impose a corresponding change in the lattice structure and that this might be detected by suitable optical methods. Attempts were made at an early date to detect such changes. To this end a mask of nichrome ribbon 5 mils wide and 1 mil thick was laid over a sample of silicon during bombardment. An optical examination of the surface showed that after bombardment in the case of helium the surface on either side of the mask was elevated whereas in the case of argon it was depressed. This result has since been confirmed by one of the authors's colleagues, Dr. F. W. Reynolds, who has found that in cases of prolonged bombardment by helium the adjacent surfaces may be elevated by as much as 225 Angstroms[7] while in the cases of prolonged bombardment by

[7] One Angstrom is 10^{-8} cm.

argon the adjacent surfaces may be depressed by as much as 130 Angstroms. Further investigation of this phenomenon is under way.

STABILITY OF BOMBARDED SURFACES

No extended test has yet been made of the stability of bombarded surfaces but results extending over more than two years are encouraging. Similarly, rectifiers for the millimeter wavelength range, mounted without the usual protective impregnation, show little or no change at the end of a year.

In a few instances bombarded surfaces have been subjected to rather severe tests with results that suggest that under normal conditions they may be even more stable than surfaces activated by more conventional methods. For example, surfaces contaminated while cutting or while cementing them to their mountings have subsequently been cleaned with solvents such as alcohol and are substantially the same before and after treatment. In other cases, they have been heated in a flame to soldering temperatures with no appreciable effects. Even in the very severe case where the bombarded piece was heated to a cherry red and the superficially oxidized layer was removed with hydrofluoric acid the effects of bombardment were still evident. There was, however, considerable reduction in the tolerable reverse voltage. There is nothing in our experience to date to suggest that bombarded surfaces treated in accordance with the simple straightforward methods outlined above, are in any wise temporary in character.

CONCLUSIONS

The experiments reported above have shown that rather pronounced changes in the electrical properties of silicon may be produced merely by bombarding the polished surface with positive ions. The ratio of forward to reverse currents, for example, which for the usual untreated silicon is seldom more than a few hundred, can be made more than 10,000. Experiments show that the effect depends to some extent on the type of ion gas used, helium being a preferred medium. The effect depends also on the volocity of the bombarding particles, the total bombarding charge and to a lesser extent on the temperature of the specimen during bombardment. Good results are obtained from bombarding potentials of 30 kv with current densities of 5 microamperes per square centimeter for periods of one or two minutes. The temperature should preferable be about 300°C.

Ordinarily the properties of silicon are materially affected by impurity

content. In the case of bombarded silicon the effect is much less. More particularly it is possible to contaminate silicon with impurities such as boron to the point where its rectifying properties are almost completely lost and by bombardment it is possible to convert the crystal into a very useful rectifier. It is possible to produce results similar to the above by exposing the crystal to radioactive polonium. Bombarded materials appear to be relatively stable.

The writer wishes to express his appreciation of the encouragement and help of Dr. G. C. Southworth in the preparation of this paper, to A. J. Mohr, Jr., for his able assistance in the experimental work, and to numerous associates in Bell Telephone Laboratories for their assistance in preparing materials and in making special tests for which the author was not adequately equipped.

Mechanical Properties of Polymers at Ultrasonic Frequencies

BY WARREN. P. MASON AND H. J. McSKIMIN

(Manuscript received October 25, 1951)

Since the mechanical properties of solid polymer materials are largely dependent on the motions that segments of the polymer chains can undergo, to understand these properties one must use measuring techniques which can determine these motions. One of the most promising methods is to measure the reaction of polymer materials to longitudinal and shear waves over a frequency spectrum wide enough to determine the relaxation frequencies due to thermal motions of the principle elements of the chain. The presence of relaxations is indicated by a dispersion in the velocity and attenuation constants of the material, or a dispersion in the characteristic impedance of the material if the attenuation is too high to allow velocity measurements. A number of different types of measuring methods are described in this paper which make possible propagation and impedance measurements not only in solid polymers, but also in liquid polymers and in solutions of polymer molecules in typical solvents.

When these techniques are applied to long chain polymers in dilute solutions, the three relaxations observed correspond to motions occurring in isolated molecules since as the dilution increases, the molecules seldom touch. The lowest relaxation corresponds to a configurational relaxation of the molecule as a whole, the highest relaxation corresponds to the twisting of the shortest segment—containing about 40 repeating units—while the intermediate relaxation corresponds to a transient entanglement of chain segments. All three types of relaxations are present in pure polymer liquids but are spread out over a frequency range due to the perturbing effect of near neighbors of adjacent chains. The high frequency shortest chain relaxation can be traced in solid polymers of the linear chain type such as polyethylene and nylon and produces rubber-like response to mechanical shocks of very short duration.

I. INTRODUCTION

The mechanical properties of solid polymer materials are largely determined by what motions, parts or segments of the polymer chains can undergo. Toughness, mechanical impact strength and ultimate elongation depend on the facility with which the polymer molecule can be displaced.

If only a small motion of the polymer chain can occur within the time of the measurement, the material has high elastic stiffness coefficients and acts similar to a rigid solid. On the other hand, if significant segments of the polymer chain can move at the frequency of measurement, the elastic stiffness is much lower and rubber-like behavior results. An intermediate case, which occurs when the significant motion of the polymer molecule is near the relaxation time at the frequency of measurement, is that of a damping material such as butyl rubber. Even "long time" qualities of plastics such as creep, stress relaxation and recovery depend on the integrated displacements of rapidly oscillating segments of the chain.

One of the most promising methods for investigating these motions is to determine the reaction of mechanical waves on the polymer materials over a wide spectrum of wavelengths, eventually going to frequencies comparable with those of thermal vibrations of significant groups or segments in the macromolecules.

If one wishes to understand the origins of these motions it is necessary to measure the molecules in the form of liquids or solutions since then the segments of the molecule are less restrained by their neighbors and can perform all the possible vibrations. Polymer liquids are also interesting in themselves as sources of damping material. To apply these results to rubbers and solid materials, one then has to measure the modifications of the polymer chain motion caused by the close approach of near neighbors, by measuring the mechanical properties of these materials.

By using different types of techniques, these processes can be applied to molecules in solution, to liquid polymers and to solid polymers. The principal types of methods used for liquids are the torsional crystal, the torsional wave propagation system and the shear wave reflectance method, all of which are described in Section II. For solids an optical method and an ultrasonic method are described in Section V. All of these methods involve displacements of 10^{-6} cm or less so that non-linear effects are negligible.

All of these methods depend on setting up shear or longitudinal waves in the medium and observing either the velocity and attenuation of the wave, or the reaction of the medium back on the properties of the transducer. If the attenuation of a wave in the medium under consideration is low enough to permit the wave parameters, i.e., the velocity and attenuation per wavelength to be determined, the relaxation of some significant part of the polymer molecule is determined by the dispersion of the wave properties which occur, as shown by Fig. 1A, in the form of an increase in velocity and a maximum in the attenuation per

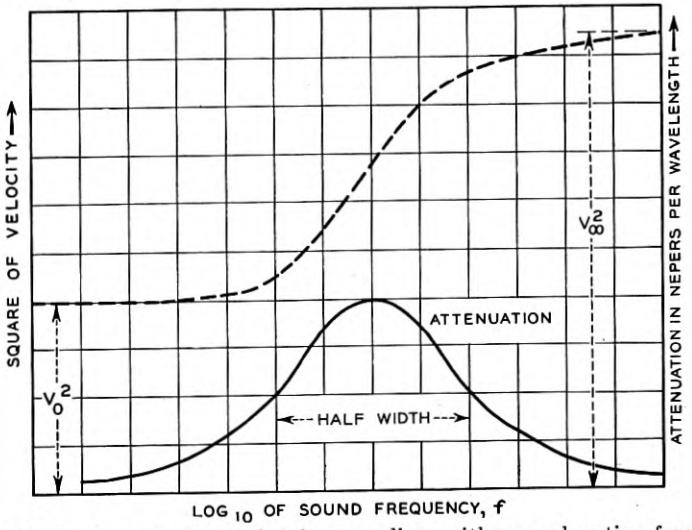


Fig. 1A—Velocity and attenuation for a medium with one relaxation frequency.

wavelength curve. If the variation in this relaxation mechanism is studied as a function of temperature and chain length, the type of segment may be determined. If, however, the attenuation of the medium is so high that its wave properties cannot be determined, some information can still be obtained by determining the loading, or mechanical impedance, that such a wave exerts on the driving crystal or transducer. If all the relaxations occur in the stress-strain relation, it can be shown that there is a reciprocal relation between the propagation constant $\Gamma = A + jB$, and the characteristic impedance per square centimeter Z_0 given by the equation

$$Z_0\Gamma = (R + jX)(A + jB) = j\omega\rho \tag{1}$$

where A is the attenuation and B the phase shift per centimeter, R the mechanical resistance and X the mechanical reactance per square centimeter, ω is 2π times the frequency and ρ the density of the medium. A typical two relaxation mechanism[1] is shown by the curves of Fig. 1B. By assuming values for the stiffness and dissipation factors and fitting a theoretical curve to the measured values, the relaxation frequency or frequencies can be determined.

[1] All the relaxation mechanisms discussed in this paper are represented in terms of equivalent parallel electric circuits in which the resistance terms represent viscosities and the inverse of capacities represent shear elastic stiffnesses. In mechanical terms these correspond to a series of Maxwell models as discussed in a paper by Baker and Heiss to be published in the next issue.

The most information about the motions of isolated polymer chains can be obtained by investigating the properties of polymer solutions. This follows from the fact that in pure polymer liquids, and in solids, the mechanical properties are mainly determined by interactions between chains on account of the close packing of the chains. If, however, one dissolves the polymer molecules in a solvent, the inter-chain and intra-chain reactions can be separated as the dilution increases. When the polymer is in the order of one percent of the solvent, the chains on the average touch very seldom and the mechanical properties of the solution are determined by the properties of single molecules. As discussed in Section III, three types of chain segment motion have been isolated, (1) a configurational relaxation of the chain as a whole, (2) a position change of the shortest segment and (3) twisting of the shortest chain segment. Above the frequency of relaxation of this chain segment the joints of the polymer molecule become frozen and the chain becomes very stiff. These shortest chain relaxations occur also in pure polymer liquids, in rubbers and in non rigid solids with linear chain segments such as polyethylene. In pure liquids a lower frequency quasi-

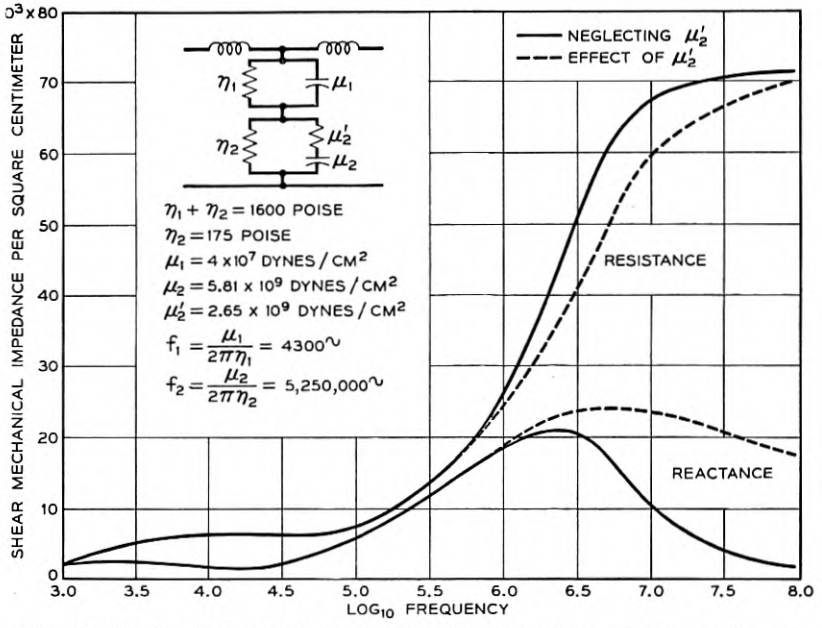


Fig. 1B—Mechanical impedance loading for a medium with two relaxation frequencies.

configurational relaxation also occurs for chain lengths greater than 60 elements but for chain lengths less than 40 elements this type of relaxation disappears. From the difference between the high frequency shear elasticities measured for polyethylene and nylon 6-6 and those measured for static pulls, it appears that there may be lower frequency relaxations in these materials as well.

II. METHODS OF MEASUREMENT FOR SOLUTIONS AND PURE POLYMER LIQUIDS

To measure the mechanical properties of such dilute solutions, shear waves have been used since for longitudinal waves the added stiffness caused by the dissolved polymers is very small compared to the stiffness of the solvent alone. The velocity and attenuation of a longitudinal wave are given by the equations

$$v = \sqrt{\frac{\lambda + 2\mu}{\rho}}; \qquad A = \frac{2\pi^2 f^2}{\rho v^3}[\chi + 2\eta] \tag{2}$$

where λ and μ are the Lamé elastic constants, f the frequency, ρ the density, v the sound velocity, χ the compressional viscosity and η the shear viscosity. Since for a one percent solution of polyisobutylene in cyclohexane the shear elasticity does not exceed 90,000 dynes/cm^2, whereas the value of λ is in the order of 2×10^{10} dynes/cm^2, it is obvious that the longitudinal velocity would have to be measured to an accuracy of 1 part in 100,000 before the presence of polymer molecules could be ascertained. Attenuation measurements give some information on the added viscosity due to the chain molecules but since longitudinal attenuations are not easily measured below 1 megacycle, the most interesting frequency range is missed.

A pure shear wave in a viscous liquid is propagated according to the equation[2]

$$v = v_0 e^{-\sqrt{\frac{\pi f \rho}{\eta}}(1+j)z} \tag{3}$$

where v is the transverse particle velocity, ρ the density, f the frequency, η the shear viscosity, $j = \sqrt{-1}$ and z the distance. For typical solvents, the attenuation is so high that wave motion cannot be measured. However the viscous wave produces an impedance loading on a crystal generating such a wave which can be measured by the change in the resonant frequency and the change in the resistance at resonance. The mechanical impedance per square centimeter caused by such a viscous

wave is equal to[2]

$$Z_0 = \sqrt{\pi f \eta \rho}\,(1 + j) = R_M + jX_M \tag{4}$$

This causes a change in resistance, and a change in frequency in a crystal generating a shear wave in the liquid equal to

$$\Delta R_E = K_1 R_M; \qquad \Delta f = -K_2 X_M \tag{5}$$

where K_1 and K_2 are constants of the crystal which can be obtained approximately from the dimensions and piezoelectric constants of the crystal but which are more accurately obtained by calibration in known liquids. The constants K_1 and K_2 vary slightly with temperature and should be calibrated over a temperature range.

The first instrument to use a vibrational method for measuring viscosity was the vibrating wire method of Phillipoff.[3] In this method a wire was vibrated in a liquid and the damping rate was used as a measure of the viscosity. Another method also applicable in the low frequency range is the transducer method of Ferry.[4] In this method wires are vibrated by electromagnetic transducers and the resistance and reactance drag on the wires are measured by the change in the electrical resistance and reactance of the transducer. From the constants of the transducer, the equivalent viscosity and stiffness of the liquid can be measured.

In the medium frequency range a torsional crystal[5] method was devised by one of the writers which has been applied in the frequency range from 10 to 150 kc. The torsional crystal is shown by Fig. 2. For these types of measurements the crystal usually is made of quartz with four electrodes of gold evaporated on the surface. Four wires are soldered on the surface and serve as supports as well as electrodes. The motion is all tangential to the surface and tests at Bell Laboratories and at the Franklin Institute,[6] where a precision study of the torsional crystal has been made, have shown no observable longitudinal waves from the crystal surface. The process of measurement consists in measuring the

[2] W. P. Mason, *Piezoelectric Crystals and Their Application to Ultrasonics*, D. Van Nostrand, 1950, p. 340.

[3] W. Phillipoff, *Physik. Zeits*, **35**, 1934, pp. 884–900.

[4] T. L. Smith, J. D. Ferry and F. W. Schemp, "Measurement of the Mechanical Properties of Polymer Solutions by Electromagnetic Transducers," J. App. Phys., **20**, No. 2, Feb. 1949, pp. 144–153.

[5] W. P. Mason, "Measurement of the Viscosity and Shear Elasticity of Liquids by Means of a Torsionally Vibrating Crystal," A.S.M.E., **69**, May 1947, pp. 359–367.

[6] P. E. Rouse, Jr., E. D. Bailey, and J. A. Minkin, "Factors Affecting the Precision of Viscosity Measurements with the Torsional Crystal," Laboratories of the Franklin Inst., Report 2048, presented to Am. Petroleum Inst., May 4, 1950.

resonant frequency and resonant resistance of the crystal in a vacuum, then introducing the solution to be measured, the change in the resonant resistance ΔR_E and the change in resonant frequency Δf are determined by an electrical bridge. Several short cuts are possible if the mechanical impedance is not too high. By measuring the capacity at a frequency considerably higher than the crystal frequency, the resistance at resonance and Δf can be obtained by changing the frequency and resistance until a balance is obtained leaving the capacity unchanged. This method has been used to measure viscosity, and a recent precision study at the Franklin Institute[6] has shown that it agrees with other methods to an accuracy of well under a per cent.

The torsional quartz crystal has been successfully used to measure liquids having a viscosity up to 10 poise, but above this viscosity the electrical resistance gets so high that it is hard to measure it since it is shunted by the much smaller reactance of the static capacitance of the crystal. A crystal of higher electromechanical coupling such as am-

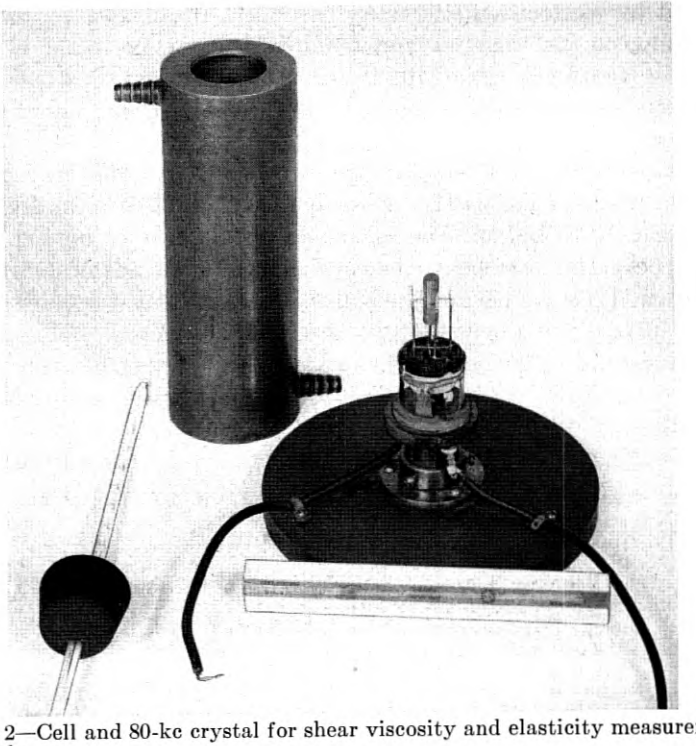


Fig. 2—Cell and 80-kc crystal for shear viscosity and elasticity measurements of liquids.

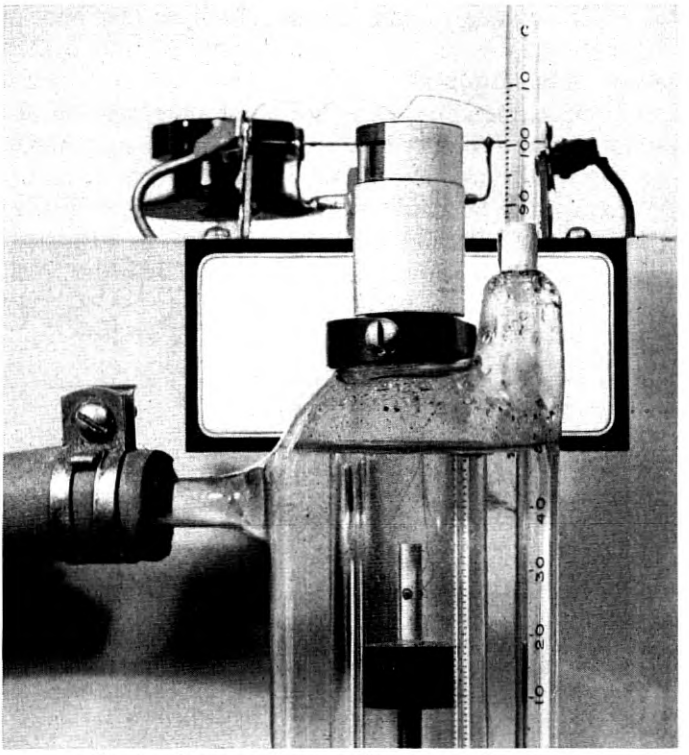


Fig. 3—Photograph of torsional crystal and rod.

monium dihydrogen phosphate (ADP) will cause the electrical resistance component to be smaller in comparison to the reactance of the static capacitance and hence can be used to measure higher viscosities. However, since wires cannot be soldered to the surface but must be glued, the crystal is much more fragile than quartz and its use has been abandoned in favor of another method which makes use of the phase and attenuation change in a torsional wave in a rod caused by the surrounding liquid whose properties are to be measured.

This method, devised by one of the writers,[7] consists in sending a short train of torsional waves, periodically repeated, down a glass or metal rod. As shown by Fig. 3, the torsional wave is generated by a torsional quartz crystal soldered or glued to the end of the rod. These waves travel to the free end of the rod and are reflected back to the crystal where they are detected, amplified, and displayed on a cathode ray

[7] This method is described by H. J. McSkimin, in a paper before the Acoustical Soc. of Am. in October, 1951.

oscilloscope. Echoes due to end to end reflections also appear, being attenuated by normal acoustic losses until they are undetectable by the time the next pulse is applied.

With only air surrounding the rod, a phase reference and amplitude reference are obtained for the first received wave (or subsequent echoes if greater sensitivity is desired). The rod is then immersed a definite length in the liquid to be measured, as shown in Fig. 4, with a resulting phase retardation and amplitude reduction. These are measured by employing the experimental circuit shown by Fig. 5. In order to allow the

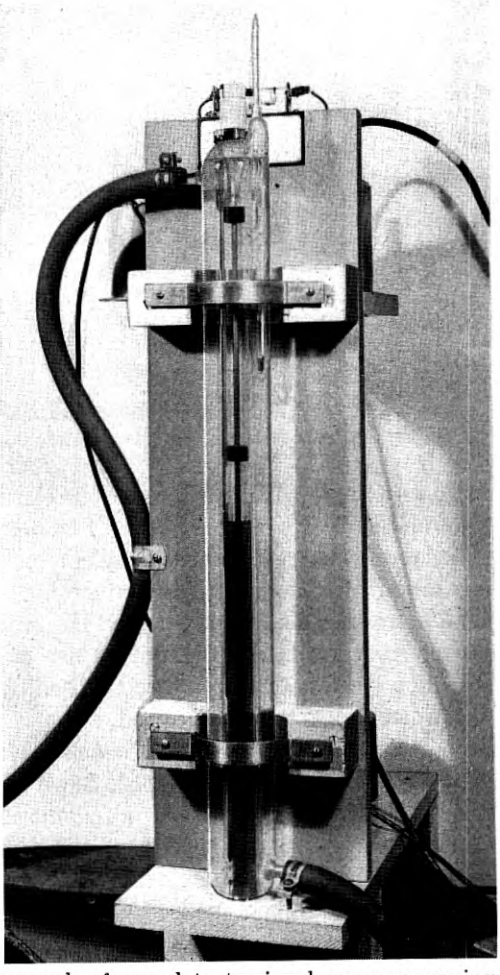


Fig. 4—Photograph of complete torsional wave measuring equipment.

use of one crystal for both receiving and transmitting, the crystal is put in the bridge circuit of Fig. 5 where a resistance and capacity are used to balance out the transmitted pulse so that it will not overload the amplifier. The relatively weak voltages generated by the incoming acoustic waves pass through directly. The gate circuit provides pulses of radio frequency voltage at repetition rates in the range of 20 to 100 per second with a synchronizing voltage supplied to the oscilloscope for the horizontal sweep. The frequency range of the device is from 20 to 200 kc. Both glass and nickel-iron rods were used, the latter having a very low frequency-temperature coefficient. With a 100-kc quartz

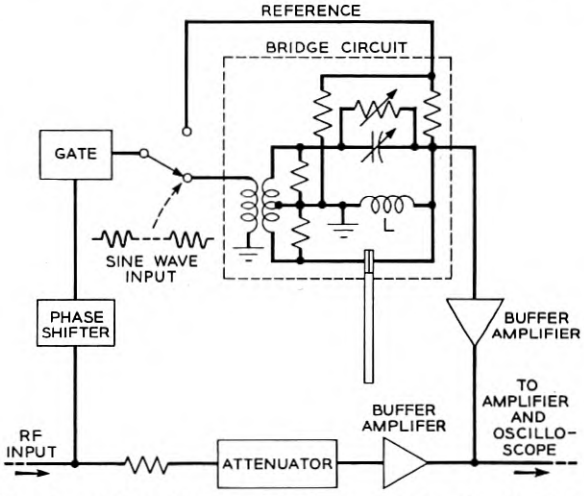


Fig. 5—Experimental pulsing circuit for measuring torsional impedance of liquids.

torsional crystal, a rod length of 21 inches and diameter of 0.2 inch were used. The entire crystal-rod assembly is placed inside a glass temperature control unit, as shown by Fig. 4, through which water can be circulated to provide temperatures in the range 0°C to 80°C. The test liquid is placed either directly into the inner bore of this water jacket, or in another tube which can be inserted from the bottom to surround the rod up to a fixed mark.

In use, both phase and attenuator settings were adjusted to balance the first received pulse against the continuous wave component passing through the attenuator. Cancellation for the duration of the pulse was visually indicated on the oscilloscope. A plot of balance phase and level is made as a function of the temperature. When the liquid is introduced an attenuation change $\overline{\Delta A}$ and a phase change $\overline{\Delta B}$ are required to

re-balance the circuit. These are measured by the amount of attenuation in nepers (1 neper = 8.68 db) and the number of radians phase shift required to re-establish a balance. An alternate method of measuring phase shift is to measure the change in frequency required to re-establish balance. If this method is used the phase shift change of the overall circuit with frequency has to be calibrated for the uncovered rod by

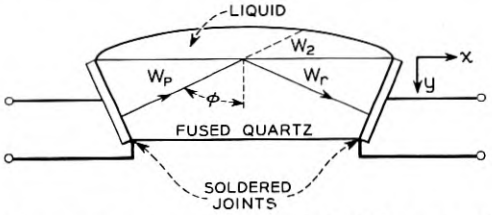


Fig. 6—High frequency shear reflection method for measuring shear impedances of liquids.

noting the frequencies for which 360° phase shifts (as measured by balance) occur in the circuit.

It is shown in the appendix that the torsional impedance of the liquid per square centimeter is given by

$$Z = \left(\frac{\rho v_0}{4}\right)\frac{a}{\ell}\left[\overline{\Delta A} + j\overline{\Delta B}\right] \tag{6}$$

where ρ is the density, and v_0 the sound velocity in the rod, a is the radius and ℓ the covered length of the rod and $\overline{\Delta A}$ and $\overline{\Delta B}$ are respectively the change in attenuation in nepers and the change in phase shift in radians to re-establish balance. If a very viscous liquid is used it may be necessary to correct for the fact that the torsional impedance may differ from the plane wave impedance as discussed in the appendix.

This device can measure liquids having dynamic viscosities from 10 poise to 1,000 poise with an accuracy of the order of 10 per cent. The frequency range covered may be from 20 to 200 kc depending on the size of the crystal used to drive the rod. Hence it supplements the torsional crystal method for very viscous liquids.

At frequencies above 500 kc, the torsional crystal becomes too small to be used practically and recourse is had to a high frequency pulsing method.[8] As shown by Fig. 6, shear waves are set up in a fused quartz

[8] W. P. Mason, W. O. Baker, H. J. McSkimin and J. H. Heiss, "Measurements of the Shear Elasticity and Viscosity of Liquids by Means of Ultrasonic Shear Waves," Phys. Rev., **75**, No. 6, March 15, 1949, pp. 936–946. See also H. T. O'Neil, "Refraction and Reflection of Plane Shear Waves in Viscoelastic Media," Phys. Rev., **75**, No. 6, March 15, 1949, pp. 928–936.

rod by means of a Y-cut or AT cut crystals soldered to a silver paste layer baked on the fused quartz surface. The particle motion of the shear wave is parallel to the large reflecting surface and hence only shear waves are reflected from this surface. These impinge on a second shear crystal which is connected to an amplifier and oscillograph. Since the attenuation in fused quartz is so low, a long series of reflected pulses appear on the oscillograph. When a liquid, whose shear properties are to be measured, is placed on the fused quartz surface, this causes a change in the amplitude and phase of the reflected wave. By using the balance method shown by Fig. 7, in which two identical fused quartz rods are used, one of which has a liquid layer and the other does not, and by using a phase shifting network and an attenuator to balance out

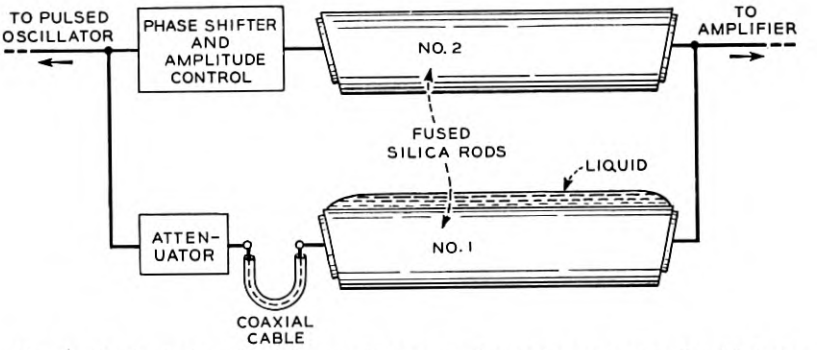


Fig. 7—Method for obtaining resistance and reactance terms for high frequency shear reflection method.

pulses, the shear impedance of the liquid can be determined. If R is the loss per reflection expressed as a current ratio, θ the change in phase angle required to rebalance the circuit and φ the angle between the wave normal and the reflecting surface, it can be shown[8] that the shear impedance of the liquid is

$$Z_M = R_M + jX_M = Z_q \cos \varphi \left[\frac{1 - R^2 + 2jR \sin \theta}{1 + R^2 + 2R \cos \theta}\right] \tag{7}$$

where Z_q is the impedance ρv for shear waves in the quartz. This is equal to

$$Z_q = 2.20 \times 3.76 \times 10^5 = 8.27 \times 10^5 \text{ mechanical ohms}$$

Since this impedance is much larger than that of the liquids that are to be measured, the sensitivity is increased by making φ large. In practice φ was taken as 80°. This method is applicable from 3 mc up to 100 mc

and complements the other methods. Fig. 8 shows a photograph of the equipment.

III. MEASUREMENTS OF POLYMERS IN SOLUTION

When such methods are applied to a polymer solution, it is found that the resistance and reactance components are no longer equal but the resistance is invariably larger than the reactance. This indicates the presence of a shear elasticity in the solution. If the molecules have a single relaxation frequency, it has been found that the shear properties of the liquid can be represented by a stress-strain equation of the type

$$T = \eta_A \frac{\partial S}{\partial t} + \frac{1}{\dfrac{1}{\eta_B \dfrac{\partial S}{\partial t}} + \dfrac{1}{\mu_B S}} \tag{8}$$

where T is the shearing stress, S the shearing strain, η_A the solvent viscosity, η_B a molecular viscosity of some particular motion of the chain which disappears when the reactance of the chain stiffness μ_B of this motion is low enough so that the motion can follow the applied shearing stress at the frequency of the measurement. When this type of mechanism is present in the liquid, it has been shown[9] that the impedance the liquid presents to the crystals is

$$Z_0 = R_M + jX_M = \frac{\rho\mu_B\eta_B^2 + j\left[\omega\rho\eta_A\eta_B^2 + \dfrac{\rho\mu_B^2}{\omega}(\eta_A + \eta_B)\right]}{\eta_B^2 + \mu_B^2/\omega} \tag{9}$$

Fig. 9 shows a plot of the resistance and reactance components of an assumed solution having a single relaxation frequency, and a viscosity 30 times the solvent viscosity. At very low frequencies, the resistance and reactance follow that of a solution, but for frequencies comparable with the relaxation frequency, the resistance becomes larger than the reactance while for very high frequencies the two come together on a line determined by the solvent viscosity. If there is more than one relaxation frequency, the resistance and reactance may coalesce for several intermediate stages. A continuous distribution would give a definite relation between the frequency dependence of resistance and reactance.

The torsional crystal and the shear wave reflection method have been applied to long chains of polyisobutylene dissolved in various sol-

[9] W. P. Mason, *Piezoelectric Crystals and Their Application to Ultrasonics*, D. van Nostrand, 1950, p. 353.

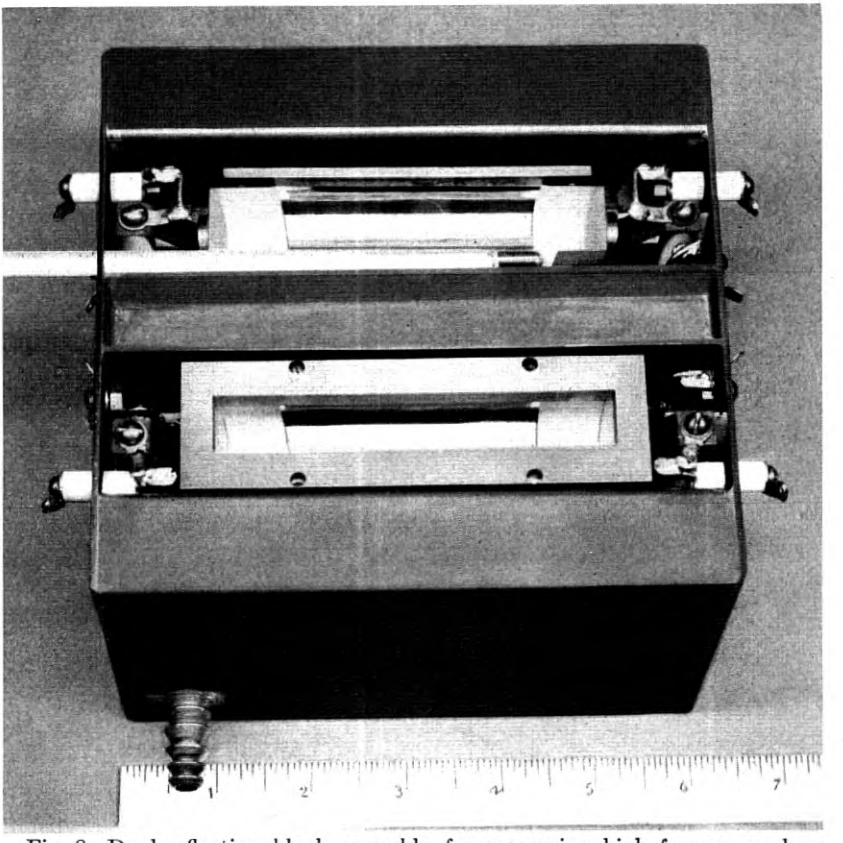
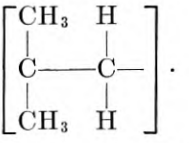


Fig. 8—Dual reflecting block assembly for measuring high frequency shear impedances of liquids.

vents with concentrations ranging from zero per cent to 10 per cent. Polyisobutylene is a polymer molecule having the chemical formula



Non-planar zigzag segments can be expected in the liquid state. Fig. 10 shows measured curves for 20 kc of the resistance and reactance for solutions of viscosity average molecular weight of 3,930,000 in cyclohexane. Four values of concentration were used and two temperatures were measured. For pure cyclohexane, the resistance and reactance

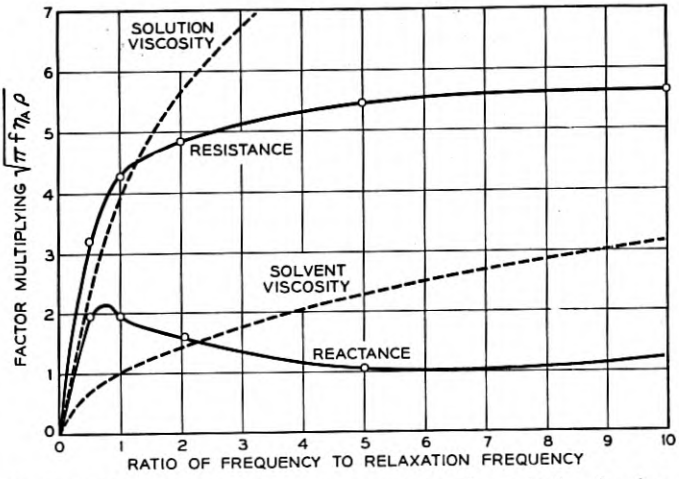


Fig. 9—Resistance and reactance components of a solution having a single relaxation frequency and a solution viscosity 30 times the solvent viscosity.

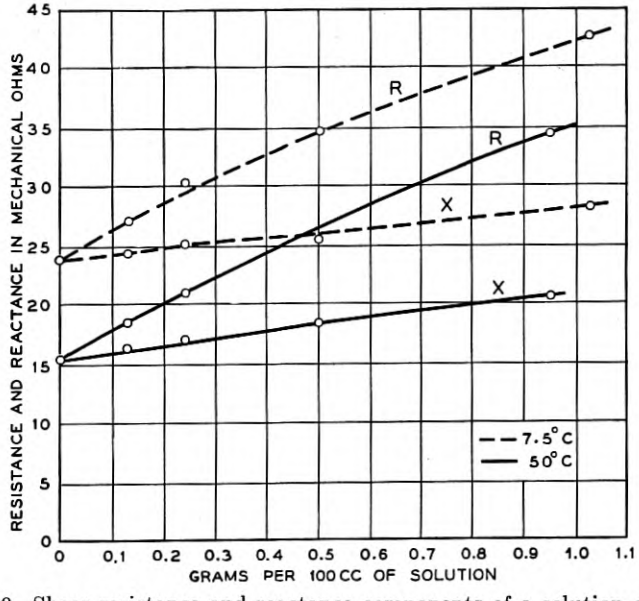


Fig. 10—Shear resistance and reactance components of a solution of polyisobutylene (molecular weight of 3,930,000) in cyclohexane plotted as a function of grams per cc of solution.

components are equal but as the percentage of polyisobutylene is increased, the resistance increases more rapidly than the reactance.

By solving equation (9) for η_A , η_B and μ_B in terms of R and X measured at one frequency and $\eta_A + \eta_B$ the solution viscosity, we find

$$\eta_A = \frac{2RX}{\omega\rho} - \frac{(R^2 - X^2)^2/\omega\rho}{\omega\rho(\eta_A + \eta_B) - 2RX}; \qquad \eta_B = (\eta_A + \eta_B) - \eta_A$$
$$\mu_B = \frac{(R^2 - X^2)\,\omega\eta_B}{\omega\rho(\eta_A + \eta_B) - 2RX} \tag{10}$$

Applying these formulae to the measured results, the curves of Fig. 11 result. The shear elasticity is directly proportional to the concentration, the viscosity η_A is only slightly larger than the solvent viscosity while the main part of the measured viscosity resides in η_B the viscosity associated with chain motion. Fig. 12 shows these three quantities for a one per cent solution measured as a function of temperature. The apparent

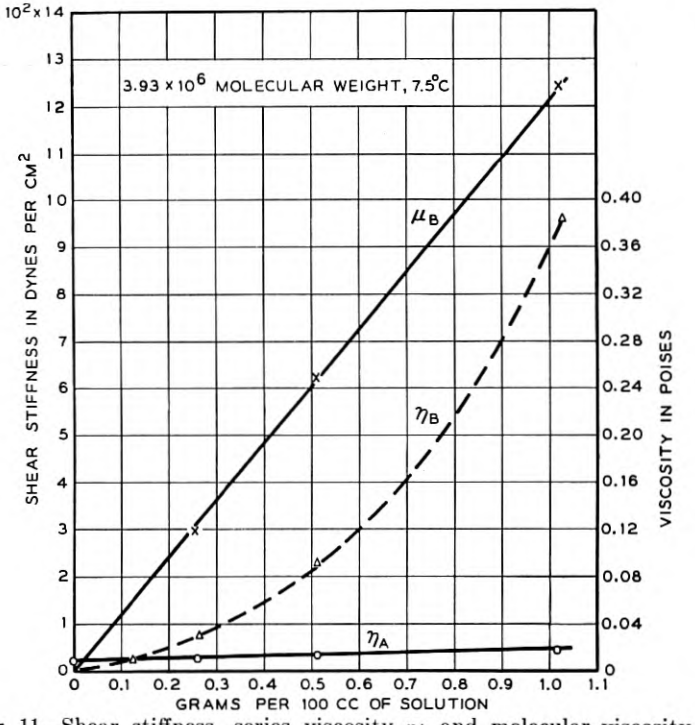


Fig. 11—Shear stiffness, series viscosity η_A and molecular viscosity η_B for polyisobutylene (molecular weight of 3,930,000) in cyclohexane plotted as a function of grams per 100 cc of solution.

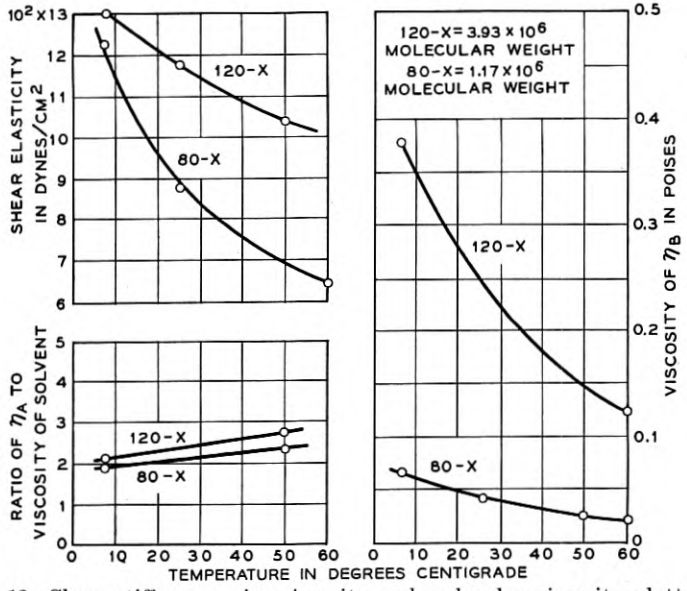


Fig. 12—Shear stiffness, series viscosity and molecular viscosity plotted as a function of temperature for two molecular weight solutions of polyisobutylene in cyclohexane.

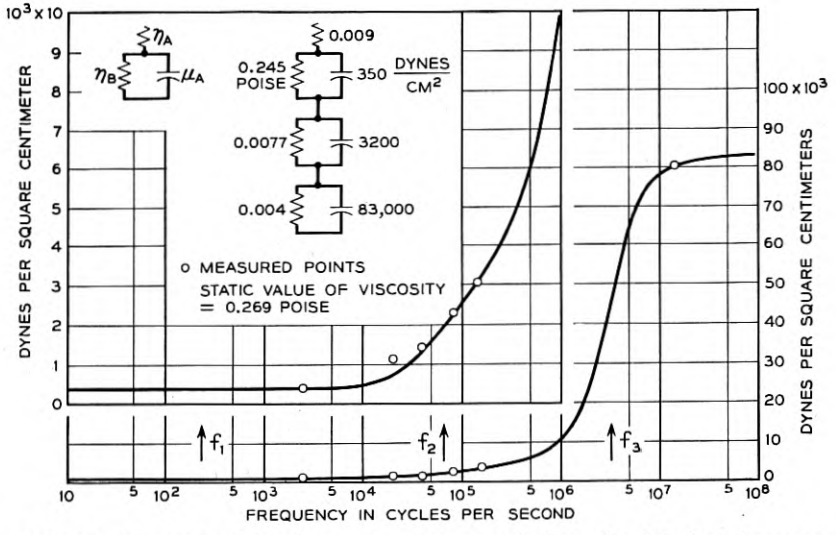


Fig. 13—Shear elasticity for a one per cent solution of polyisobutylene as a function of frequency for 25°C.

stiffness decreases with increase in temperature for a single frequency measurement. However, when measurements were made at 20, 40, 80 and 150 kc it was found that the elasticity was a function of the frequency, which indicates the presence of more than one relaxation and complicates the determination of the temperature relationships. Fig. 13 shows measurements of the shear elasticity over a frequency range from 2.5 kc[10] up to 14 megacycles for 25°C. There is a gradual rise up to about 300 kilocycles after which there is a sharp break to a stiffness of about 90,000 dynes/cm^2 for a 1 per cent solution of the 3,930,000 molecular weight polymer in cyclohexane. If one analyzes the frequency variation of the elasticity he finds that it can be fitted by three relaxation frequencies, one having a frequency of 230 cycles, one around 66,000 cycles and one around 4 megacycles. A possibility exists for a fourth relaxation. The lowest relaxation is thought to be a configurational relaxation of all the elements of the chain. The highest one appears to be a relaxation of the twisting motion of the smallest segment of the chain such as a Kuhn segment. The intermediate relaxation appears to be due to the motion of the ends of the smallest chain segment from one position of entanglement to an adjacent position. This interpretation is based partly on the fact that the associated viscosity of the motion is very similar to that for the relaxation of the twisting motion of the smallest chain segment and partly from data presented in the next section on pure polymer liquids which shows a lower frequency relaxation agreeing in frequency assymptotically with this one, which involves chain motions of approximately 30 to 40 chain elements. Temperature variations of these elastic components show that the lowest relaxation mechanism has a stiffness that increases slightly with temperature in agreement with the kinetic theory of elasticity. The corresponding viscosity (η_2) which comprises most of the viscosity for a solution, when plotted against the reciprocal of the temperature, as shown by Fig. 14, indicates an activation energy of 3.9 kilocalories per mole which is slightly higher than that of the solvent cyclohexane alone, which is about 3.2 kilocalories per mole. This difference of 0.7 kilocalories presumably represents the added energy required to bend the chain in its configurational motion. Measurements with another chain length of 1.18×10^6 molecular weight showed that the stiffness of the lowest (configurational) relaxation decreased from 310 to 160 indicating that the stiffness of this motion is approximately proportional to the square root

[10] The lowest frequency, 2.5 kc, was measured by means of a quartz crystal tuning fork which will be described in another paper. This instrument makes possible the direct measurement of configurational elasticities.

of the molecular weight. The viscosity decreased by a factor of 6.25 and consequently the relaxation frequency increases from 230 to 660 cycles.

The second "entanglement" relaxation has a stiffness of about 3100 dynes/cm^2 for a 1 gram per cc solution of the 3.93×10^6 molecular weight solution and about 2650 dynes/cm^2 for the 1.18×10^6 molecular weight solution. The variation with temperature, if any, is small. The corresponding viscosities η_3 for the two solutions are nearly equal as shown by Fig. 14, and have an activation energy of 4.25 kilocalories per mole. The final high frequency "short segment" relaxation has a high stiffness of 83,000 dynes/cm^2 for a 1 per cent solution of 3.93×10^6 molecular weight. The corresponding viscosities for the two solutions shown by Fig. 14 have nearly identical values and an activation energy of 4.25 kilocalories per mole, i.e. very closely equal to the "entanglement" relaxation viscosity.

These upper two relaxations persist in pure liquid polymers as discussed in the next section, although they are spread out over a small range of relaxation time values. The highest one can be traced in measurements of mechanical properties of solid plastics such as polyethylene and nylon which indicates that these materials should have rubber like

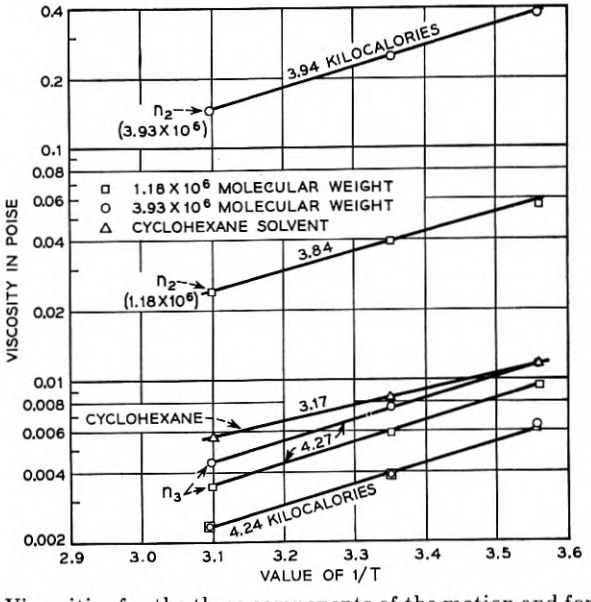


Fig. 14—Viscosities for the three components of the motion and for the solvent plotted against the inverse of the absolute temperature.

TABLE I

Liquid	Average Molecular Weight	Density				Melt Viscosity Poise				No. of Chain Segments
		15°C	25°C	35°C	50°C	15°C	25°C	35°C	50°C	
A	904	0.83	0.826	0.821	0.814	1.10	0.67	0.33	0.16	16.2
B	1,220	0.846	0.842	0.837	0.83	2.64	1.40	0.65	0.297	22.2
C	1,590	0.856	0.849	0.846	0.840	8.4	4.0	2.1	0.87	28.4
D	2,450	0.872	0.866	0.863	0.857	100	38	15.9	4.9	44.0
E	3,520	0.881	0.877	0.872	0.866	580	216	95	30	63.0
F	4,550	0.89	0.886	0.882	0.875	1,960	737	320	98	81.4
G	5,590	0.897	0.892	0.887	0.882	4,600	1,840	740	220	100
H	10,380	0.912	0.908	0.904	0.896	11,500	4,600	2,050	625	186

Measurement at 24 megacycles $\frac{\mu \text{ dynes}}{\text{cm}^2}$; η in poise

Liquid	15°C		25°C		35°C		50°C	
	μ	η	μ	η	μ	η	μ	η
A	0.482×10^9	1.07	0.4×10^9	0.68	0.35×10^9	0.39	0.1×10^9	0.13
B	1.08×10^9	2.64	0.8×10^9	1.39	0.58×10^9	0.7	0.17×10^9	0.293
C	1.65×10^9	5.12	1.2×10^9	2.57	0.885×10^9	1.99	0.27×10^9	0.881
D	2.71×10^9	15.7	2.0×10^9	8.1	1.67×10^9	4.15	0.98×10^9	1.77
E	3.81×10^9	40	3.0×10^9	20	2.17×10^9	10.6	1.2×10^9	4.62
F	5.48×10^9	73	4.22×10^9	38.3	2.8×10^9	20.2	1.6×10^9	8.8
G	5.9×10^9	107	4.78×10^9	51	3.74×10^9	29.5	2.5×10^9	13.2
H	6.9×10^9	119	5.55×10^9	65.8	4.22×10^9	31.5	2.9×10^9	14.8
Measurements at 14 megacycles								
A	0.38×10^9	0.975	0.3×10^9	0.6	0.22×10^9	0.35	0.12×10^9	0.15
B	0.475×10^9	2.53	0.35×10^9	1.4	0.25×10^9	0.74	0.18×10^9	0.53
C	0.68×10^9	3.86	0.61×10^9	3.4	0.3×10^9	1.62	0.25×10^9	0.56
D	2.32×10^9	16.3	1.7×10^9	10	0.94×10^9	4.75	0.39×10^9	1.86
E	3.8×10^9	56.2	2.8×10^9	24.25	2.0×10^9	12.3	0.855×10^9	5.7
F	4.75×10^9	90.4	3.6×10^9	48.3	2.65×10^9	26.6	1.47×10^9	10.8
G	6.03×10^9	136.5	4.6×10^9	81.4	3.24×10^9	39.6	2.0×10^9	15
H	6.64×10^9	160	5.3×10^9	93.5	3.98×10^9	54.2	2.3×10^9	18.5
Measurement at 4.5 megacycles								
A	0.34×10^9	1.18	0.19×10^9	0.64	0.18×10^9	0.34	0.09×10^9	0.17
B	0.55×10^9	2.65	0.21×10^9	1.33	0.2×10^9	0.68	0.1×10^9	0.20
C								
D	1.57×10^9	24	0.96×10^9	11.7	0.72×10^9	5.52	0.34×10^9	2.1
E	2.79×10^9	76	1.9×10^9	37.4	1.43×10^9	16.5	0.9×10^9	5.15
F	3.57×10^9	124	2.5×10^9	61.2	1.86×10^9	31	1.04×10^9	12.8
G	4.4×10^9	176	3.4×10^9	98.5	2.6×10^9	54	1.6×10^9	22.6
H	5.65×10^9	186	4.3×10^9	124	2.8×10^9	76	1.8×10^9	28.8

response to mechanical shocks of very short duration. The lowest frequency configurational relaxation is spread over a wide spectrum of relaxation times in pure liquids.

Measurements[11] of these and other chains in various solvents have also been made and the results are discussed, from a chemical point of view, in a companion paper by W. O. Baker and J. H. Heiss. It is shown that the stiffnesses vary with the polymer chain and the solvent used.

IV. MEASUREMENTS OF PURE LIQUID POLYMERS

A. Shear Wave Measurements in Liquid Polymers

Similar shear wave measurements have been made for pure polyisobutylene liquids of molecular weights from 904 to 10,380 (i.e. from 16 chain elements to 186 chain elements), by the techniques described in Section II. Some of these results have been discussed in reference (8) but the much more comprehensive measurements made since require some revisions of the original conclusions.

The easiest data to interpret are the high-frequency data obtained by the shear wave reflectance method. The data of Table I give measurements of 8 liquids varying in average molecular weight from 900 to 10,380, at three frequencies and four temperatures. If we plot for example the Maxwell shear stiffness and viscosity for the three frequencies and for 25°C as a function of the number of chain elements (here a chain element is taken as two adjacent carbon atoms one of which has two methyl groups attached and the other two hydrogens) the 4.5-mc measurements are shown by the triangles of Fig. 15. The 14 megacycle measurements are shown by the circles and the 24-mc measurements by the squares.

An attempt was made to fit these measurements with a two relaxation mechanism shown by the figure with two stiffnesses which are taken to be independent of the molecular weight and equal respectively to 1.2×10^8 dynes/cm^2 and 6×10^9 dynes/cm^2. The best fit is obtained by taking the two viscosities η_1 and η_2 equal and these are adjusted for the different molecular weights in such a manner as to best fit the experimental curve. A fair agreement is obtained except for the range from 60 to 90 chain elements where the two relaxation model gives too rapid an increase of stiffness with increase in the number of chain elements and at the high molecular weight viscosity range where the viscosity shows a dispersion in values but the model does not. The sum of the two vis-

[11] These results on the mechanical impedance of long chain molecules in solvents have been presented at the XIIth International Congress of Pure and Applied Chemistry by W. O. Baker, W. P. Mason and J. H. Heiss, Sept. 13, 1951.

cosities η_1 and η_2 assumed as a function of molecular weight is shown by Fig. 16. The log of the viscosity starts proportional to the molecular weight but above a molecular weight of 2,400 the increase is very slow and becomes asymptotic to a value of 240 poises. An equation which fits the increase in viscosity with molecular weight is

$$\eta_D = Ke^{11.8 \tanh Z/2370}$$

where Z is the molecular weight. The solid line shows a plot of this curve and the circles are the assumed values to obtain a best fit to the meas-

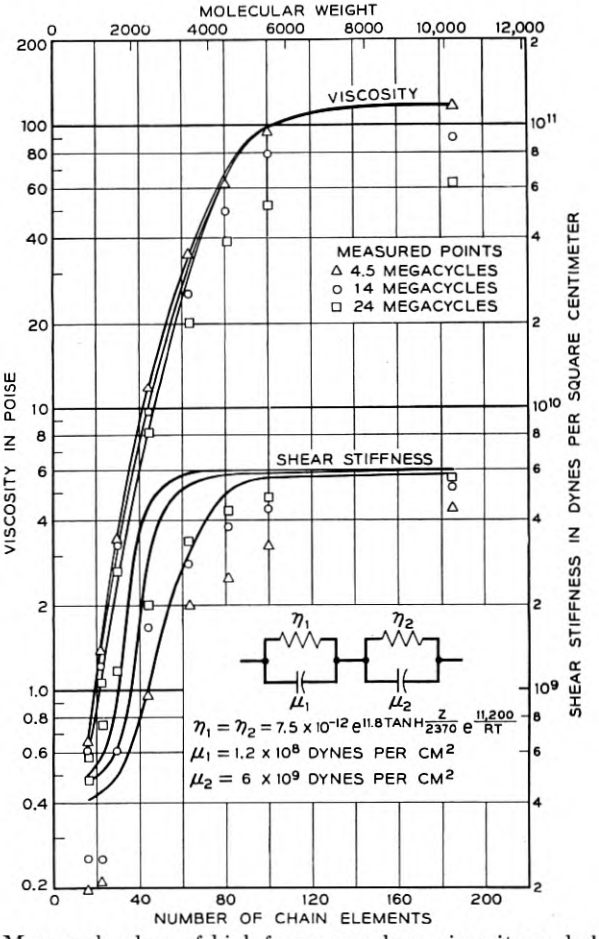


Fig. 15—Measured values of high frequency shear viscosity and elasticity for 25°C and three frequencies plotted against molecular weight. Solid lines are best fit obtained by a two relaxation mechanism having the element values shown by the figure.

ured values. This equation indicates that when $Z = 2{,}370$ or 43 chain elements the viscosity increases only a small amount more by a chain articulation effect and hence in this high frequency range we are dealing with a chain length of about 40 elements or 80 carbon atoms. This is checked also by a comparison of the static and dynamic viscosity. The total dynamic viscosity due to the two relaxation mechanisms compared to the static viscosity does not differ markedly until the number of chain elements is more than 40. Above this value other motions than that of the shortest chain segment can take place and can add to the dynamic viscosity. The static viscosity fits an equation of the same sort, but the indicated chain length for the viscous motion is about $\frac{5}{3}$ times that of the shortest segment.

When a similar process is carried out over the temperature range the equations of Fig. 16 are obtained. The static viscosity has an activation energy of 16 kilocalories per mole, while the dynamic viscosity has an activation energy of about 11.2 kilocalories per mole.

The relaxation frequencies for the two components are plotted as a function of the number of chain segments by the solid lines of Fig. 17.

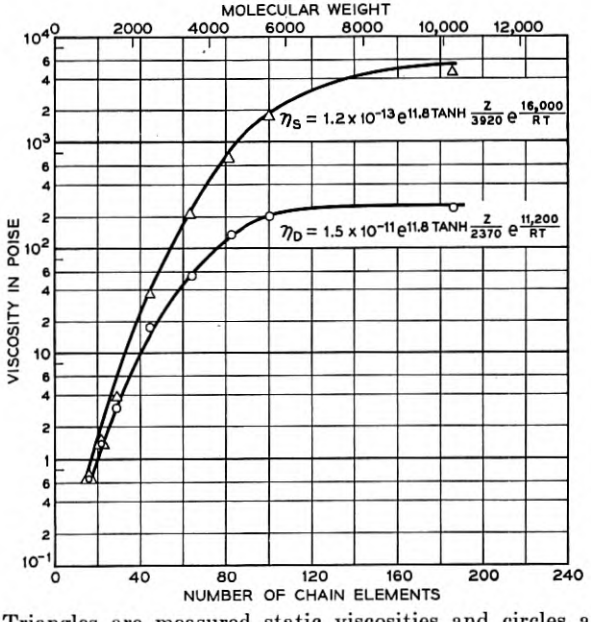


Fig. 16—Triangles are measured static viscosities and circles are dynamic viscosities plotted as a function of molecular weight. Solid lines are a plot of the equations given for static and dynamic viscosities.

For long chain segments the values become asymptotic to 8×10^6 cycles and 160,000 cycles which are not far from the two highest relaxation frequencies obtained from the solution measurements of Section III. Hence it appears likely that these relaxations are due to the "entanglement" motion and the twisting motion of the shortest chain segment. The increased activation energy is due to the fact that more energy has to be applied to the chain segment to break it loose from its equilibrium position when it is surrounded by adjacent polyisobutylene molecules than when it is surrounded by cyclohexane molecules. The stiffness of the chain is due more to the slope of the potential well than to any intrinsic chain stiffness as is shown by Fig. 18, which shows the two stiffnesses as a function of temperature. These values are obtained by fitting the 15°, 35° and 50° data in a similar manner to that used for the 25°

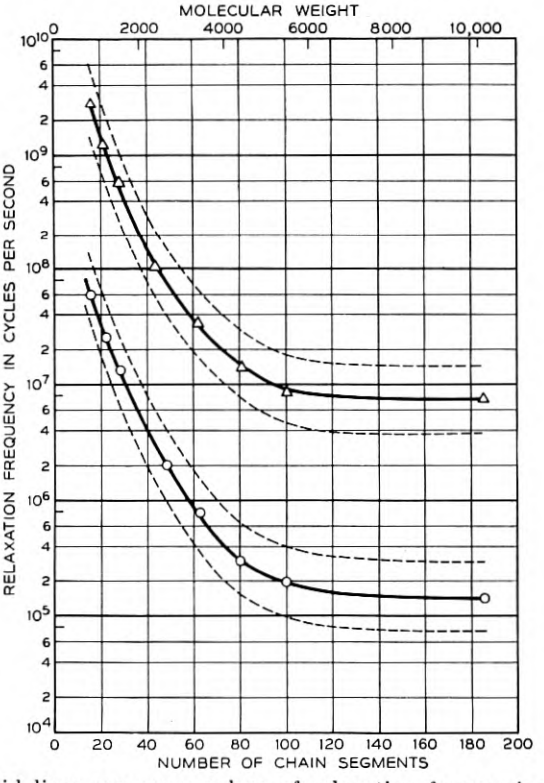


Fig. 17—Solid lines are mean values of relaxation frequencies for the two mechanisms plotted against molecular weight. Dotted lines indicate limits of regions assumed to obtain a better fit to the measured values.

data, and the activation energy of 11.2 kilocalories per mole is obtained in a similar manner.

Due to the closeness of the surrounding polyisobutylene molecules, one would expect that the relaxation frequencies would not have discrete values but would be spread about the center value in some sort of a Gaussian distribution. If we approximate this by representing each region by two relaxations, one-half, and the other twice the frequency of the mean value, as shown by the dotted lines of Fig. 17, the agreement with the measured values of Fig. 15 is considerably better as shown by Fig. 19. A wider distribution yet is indicated.

For molecular weights greater than 2,000 the shortest chain segment viscosity begins to diverge from the static viscosity indicating that there are other relaxations for these longer chains. Some data for the three longest chain polymers, F, G and H have been obtained by the torsional rod method and the results are given in Table II. These data are plotted on Fig. 20 as a ratio of dynamic to static viscosity plotted

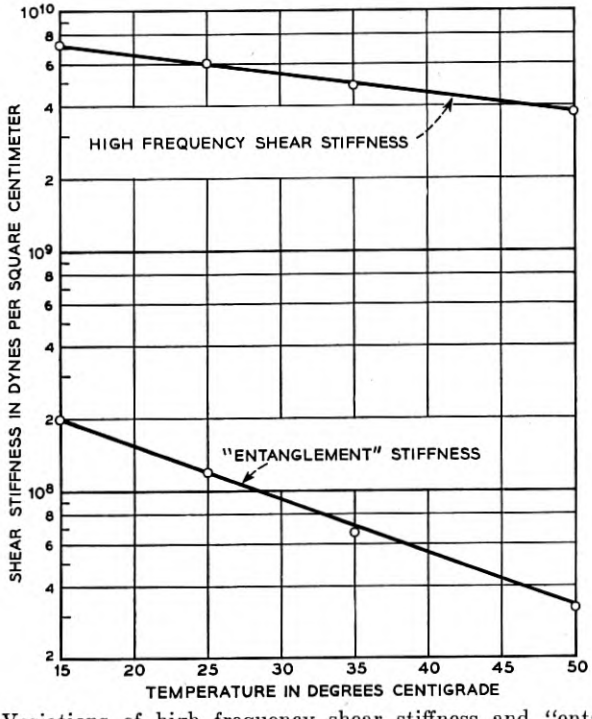


Fig. 18—Variations of high frequency shear stiffness and "entanglement" stiffness plotted as a function of the temperature.

against the frequency times the static viscosity. All the viscosity data can be represented within the experimental error by a single curve, but the stiffness curves appear to require different curves for different temperatures. On analyzing the data in terms of a distribution of relaxation frequencies, a single curve for all temperatures could be obtained if the stiffness of each mechanism were independent of the temperature and the relaxation frequency were inversely proportional to the static viscosity, i.e., had an activation energy variation equal to that for the static viscosity for each relaxation mechanism. This condition holds

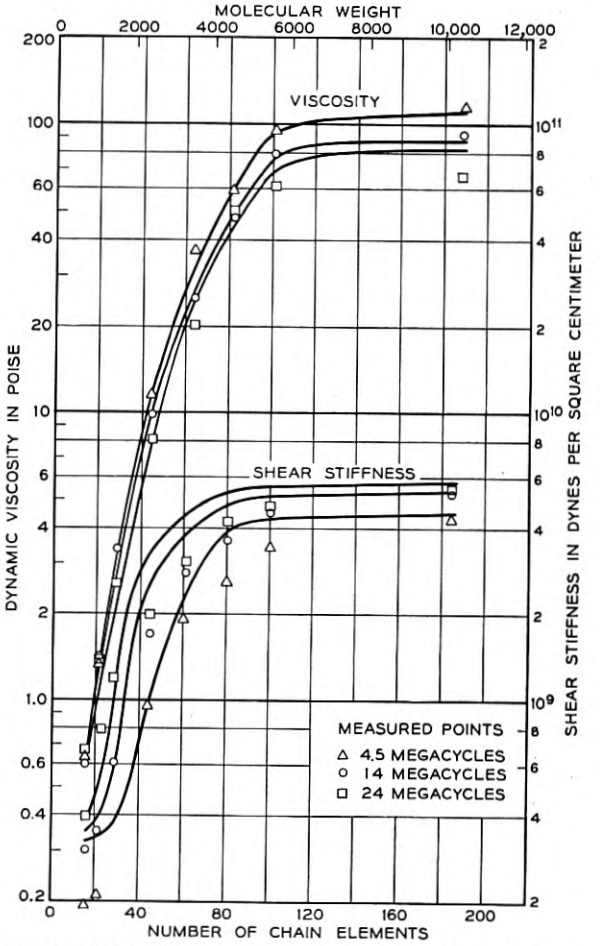


Fig. 19—Curves showing better fit to experimental data obtained by assuming the relaxation regions shown by Fig. 17.

TABLE II

Measurement for Polymer F—Molecular weight = 4550

Frequency Kilocycles	15°C		25°C		35°C		45°C		55°C		65°C	
	μ	η	μ	η	μ	η	μ	η	μ	η	μ	η
25	2.9×10^8	600	1.45×10^8	270	0.78×10^8	145	0.44×10^8	77	0.26×10^8	43	0.17×10^8	24
40	3.8×10^8	550	1.7×10^8	250	0.86×10^8	125	0.47×10^8	66	0.28×10^8	38	0.20×10^8	22
52	4.0×10^8	470	1.9×10^8	220	0.78×10^8	110	0.54×10^8	59	0.33×10^8	34	0.21×10^8	20
140	5.8×10^8	380	3.0×10^8	180	1.5×10^8	91	0.71×10^8	48	0.44×10^8	26	0.28×10^8	17

Polymer G—Molecular weight = 5590

Frequency Kilocycles	24.5°C		65°C	
	μ	η	μ	η
27.5	2.0×10^8	510	1.9×10^7	38.4
40.96	2.95×10^8	480		
54.23	3.60×10^8	456	2.93×10^7	36.1

Polymer H—Molecular weight = 10,380

Frequency Kilocycles	55°C		67°C	
	μ	η	μ	η
32.2	3.9×10^7	150	2.9×10^7	94
40.5	4.8×10^7	137	3.5×10^7	86
47.0	5.3×10^7	126	3.9×10^7	79

quite well for frequencies much lower than the relaxation frequencies of the smallest chain segment, but as the frequency approaches these relaxation frequencies, the stiffness of these polymers increases as the temperature decreases.

A fair approximation to these measured values is obtained by assuming one more "configurational" relaxation frequency in addition to the two smallest segment relaxations discussed previously. Fig. 21 shows calculations of the ratio of dynamic to static viscosity and the shear stiffness for 65°C and 25°C. The lowest relaxation frequency is assumed to have a stiffness of 6.3×10^6 dynes/cm^2 and a viscosity of 20 poises at 65°C. For 25°C the stiffness of 2×10^7 dynes/cm^2 is assumed and an activation energy of 17.3 kilocalories gives the component a viscosity of 607 poises at 25°C, and a relaxation frequency of 5,250 cycles. The average value of 16 kilocalories for the static viscosity is a result of the sum of the variation due to the two components. Although the agreement can be improved by assuming distributions of relaxation frequencies centered around these three primary frequencies, there does not seem to be much doubt of the existence of these primary relaxation

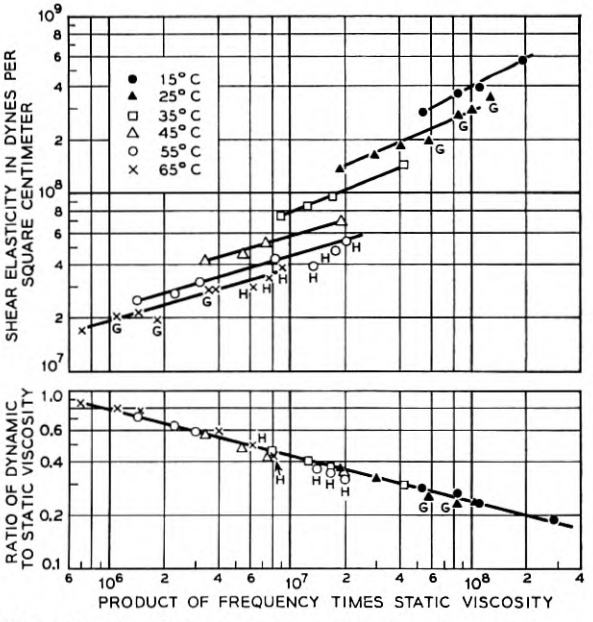


Fig. 20—Plot of ratio of dynamic to static viscosity and the corresponding intermediate frequency shear stiffnesses as a function of temperature and product of frequency times static viscosity.

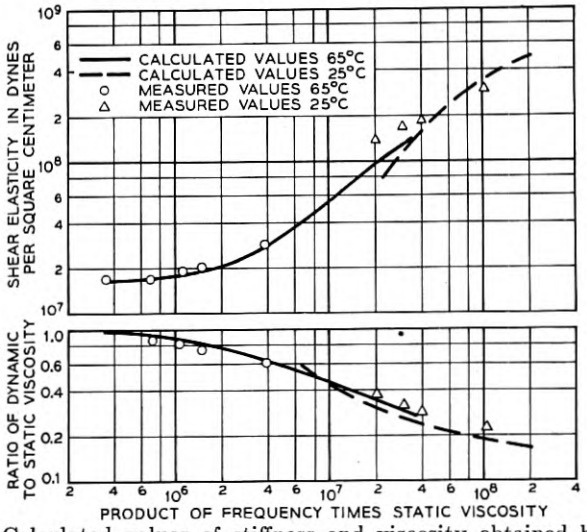
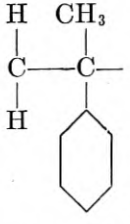


Fig. 21—Calculated values of stiffness and viscosity obtained by adding a single "configurational" relaxation frequency to the two short chain relaxations obtained from high frequency measurements.

mechanisms which show up as discrete relaxations in long chain molecules in solution.

Some measurements have also been made to determine the effect of chemical substitutions in the polymer chains. In a previous paper,* the high frequency properties of poly-α-methyl styrene were discussed. This material has the polyisobutylene chain but with one methyl replaced by a phenyl so that its chain becomes

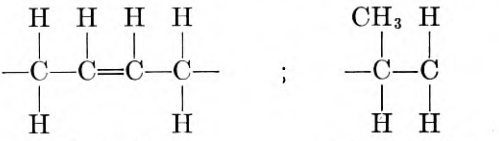


For low molecular weights this material is liquid. The shear stiffness of this liquid is somewhat higher than for polyisobutylene but has about the same change with temperature. The variation of the high-frequency viscosity, however, is much larger for poly-α-methyl styrene than for polyisobutylene, and corresponds to an activation energy of 23.6 kilocalories. The relaxation region for the shortest chain motion is much

* See footnote on page 132.

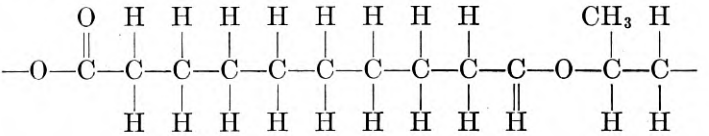
narrower than in polyisobutylene, which correlates with the smaller steric hindrance.

Measurements have also been made in the 70-kc to 140-kc range for two non polar liquids, polybutadiene and polypropylene and one polar liquid, polypropylene sebacate. The first two have the formulae



polybutadiene polypropylene

while the polar liquid, polypropylene sebacate, has the formulae



The measured results are given by Table III. These data are plotted on Fig. 22 as a function of the product of frequency times static viscosity. For comparison the data for polyisobutylene polymer F is also plotted. By extrapolating to low values of the product-frequency times static viscosity it is seen that the low frequency "quasi-configurational" stiffnesses of these liquids run from 10^6 to 1.5×10^7 dynes per square centimeter. Polyisobutylene has the greatest stiffness for any value of frequency times viscosity while polybutadiene has the least. This reflects the greatest steric hindrance of polyisobutylene and the smallest for polybutadiene which has only hydrogens connected to the carbon chain atoms. Another consequence of the larger steric hindrance of the CH_3 groups of polyisobutylene is that the viscosity associated with the shortest chain segment motion is largest for polyisobutylene and smallest for polybutadiene as can be seen from the ratio of dynamic to static viscosities for high values of frequency times static viscosity.

The activation energy for static viscosity are for polypropylene, polypropylene sebacate and polybutadiene respectively 21.2, 12 and 8 kilocalories compared to 16 kilocalories per mole for polyisobutylene. The high-frequency activation energies as determined by the dynamic measurements are respectively 11.8, 4.6 and 1.4 kilocalories for polypropylene, polypropylene sebacate and polybutadiene. The differences between the activation energy for static flow and that for dynamic flow are respectively 9.6, 7.4 and 6.6 kilocalories, which values are all higher than

TABLE III

Temp. °C	Density ρ	Poise Static Viscosity η_s	76.5 Kc Measurements				142.6 Kc Measurements			
			Maxwell		η_D/η_s	$f\eta_s$	Maxwell		η_D/η_s	$f\eta_s$
			η_D poise	μ dynes/cm²			η_D poise	μ dynes/cm²		
Polybutadiene										
10	0.877	740	25.5	7.83×10^6	0.0355	5.66×10^7	14.65	9.96×10^6	0.0196	10.55×10^7
20	0.873	450	21.1	6.38×10^6	0.0469	3.44×10^7	13.45	7.91×10^6	0.0299	6.41×10^7
30	0.869	288	17.45	5.3×10^6	0.0605	2.21×10^7	12.4	6.67×10^6	0.043	4.11×10^7
40	0.865	189	15.3	4.67×10^6	0.081	1.44×10^7	11.05	5.96×10^6	0.0583	2.69×10^7
50	0.861	128	11.9	3.92×10^6	0.093	0.98×10^7	9.16	5.21×10^6	0.0715	1.82×10^7
60	0.857	88	9.85	3.55×10^6	0.112	0.67×10^7	7.51	4.54×10^6	0.0855	1.25×10^7
Polypropylene										
			76.5 Kc Measurements				142.5 Kc Measurements			
45	0.849	30,000	380	3.56×10^8	0.01265	2.3×10^9	354.5	6.69×10^8	0.0118	4.26×10^9
55	0.842	10,900	160.2	1.4×10^8	0.0153	0.832×10^9	167	3.18×10^8	0.0153	1.55×10^9
65	0.836	4,200	89.8	0.786×10^8	0.0214	0.321×10^9	79.1	1.46×10^8	0.0188	0.6×10^9
75	0.830	1,600	50.8	0.444×10^8	0.0318	0.128×10^9	46.6	0.998×10^8	0.0292	0.23×10^9
85	0.824	700	31.2	0.264×10^8	0.0446	0.053×10^9	28.4	0.51×10^8	0.0405	0.10×10^9
Polypropylene Sebacate										
			77.1 Kc Measurements				141.5 Kc Measurements			
5	1.076	9,200	95.2	4.88×10^7	0.0103	7.1×10^8	74.6	7.45×10^7	0.0081	1.3×10^9
15	1.068	4,200	67.6	2.94×10^7	0.0161	3.24×10^8	54.9	4.28×10^7	0.0131	0.595×10^9
25	1.060	2,060	51.4	2.01×10^7	0.0249	1.6×10^8	41.3	2.82×10^7	0.02	0.292×10^9
35	1.051	1,050	43.1	1.54×10^7	0.0411	0.81×10^8	33.2	1.97×10^7	0.0316	0.148×10^9
45	1.042	580	38.8	1.24×10^7	0.0669	0.45×10^8	26.9	1.69×10^7	0.0464	0.082×10^9
55	1.034	320	32.8	1.01×10^7	0.1023	0.25×10^8	24.2	1.34×10^7	0.0756	0.045×10^9
65	1.026	183	26.8	0.88×10^7	0.147	0.14×10^8	22.9	1.25×10^7	0.125	0.026×10^9

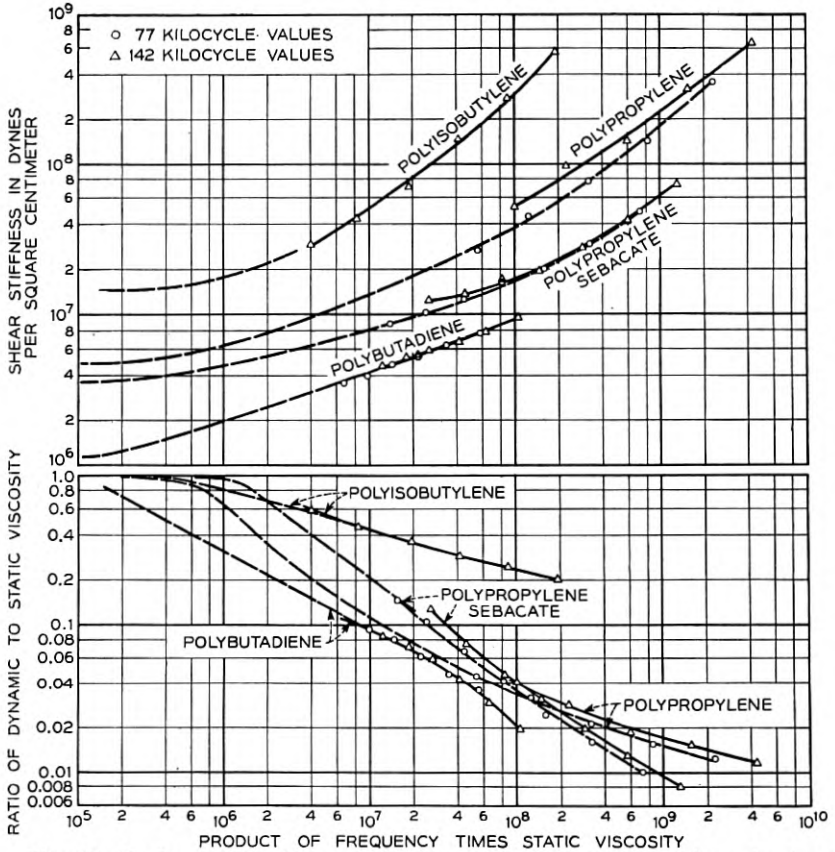


Fig. 22—Ratio of dynamic to static viscosity and the shear stiffness for four polymer liquids plotted against product of frequency and static viscosity.

the 4.8 kilocalories for polyisobutylene. This presumably indicates that there is more of a difference between the viscosity flow segment and the shortest chain segment in these materials than in polyisobutylene. Since no measurements are available over a range of molecular weights, no direct evidence has been obtained for the various chain lengths.

B. Longitudinal Wave Measurements in Liquid Polymers

Since the increase in shear elasticity for the highest relaxation frequency is so large, it should also appear in longitudinal wave measurements. Fig. 23 shows a calculation for the 5590 molecular weight liquid of the longitudinal velocity assuming that the Lamé λ elastic constant is independent of frequency and that all the variation occurs in the shear

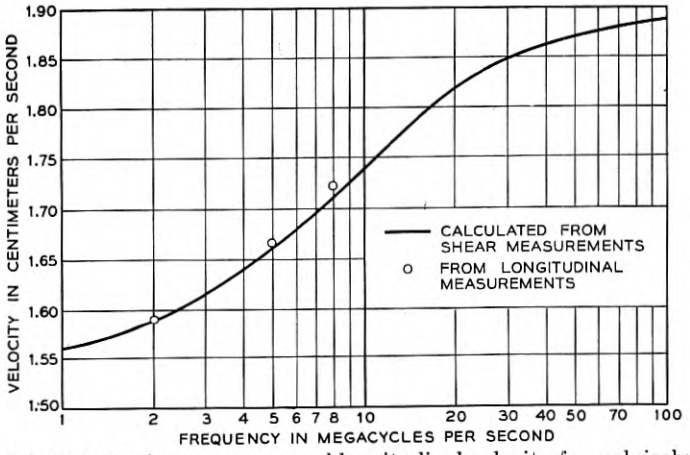


Fig. 23—Relation between measured longitudinal velocity for polyisobutylene of molecular weight 5590 and that calculated from shear stiffness measurements assuming the Lamé λ elastic constant is independent of frequency.

constant as determined by the shear measurements. The points are velocities measured for longitudinal waves and as can be seen, the measurements agree closely with the calculated values. A slightly better agreement would be obtained if λ increased by a small amount as the frequency increased. As discussed in the next section there is some experimental evidence for an increase in λ in nylon 6-6 and in polyethylene.

The question also arises as to how much of the attenuation is due to shear mechanisms and how much due to pure compressional effects. From longitudinal velocity and attenuation measurements at 30°C for the polymers E, F and G of Table I, the values of $\lambda + 2\mu$ and $\chi + 2\eta$ can be determined and are shown by Table IV. The values of μ and η can be obtained from Table I by interpolation and are given in columns

TABLE IV

Polymer	$\lambda + 2\mu$ dynes/cm²	$\chi + 2\eta$ poise	μ dynes/cm²	η poise	λ dynes/cm²	χ
			5 megacycles			
E	1.92×10^{10}	62	0.17×10^{10}	26	1.60×10^{10}	10
F	2.26	107	0.23×10^{10}	46	1.70×10^{10}	15
G	2.38	170	0.31×10^{10}	75	1.76×10^{10}	20
			8 megacycles			
E	2.01×10^{10}	50	0.20×10^{10}	22	1.61×10^{10}	6
F	2.26	93	0.27×10^{10}	41	1.72×10^{10}	11
G	2.56	155	0.34×10^{10}	69	1.88×10^{10}	17

4 and 5. Columns 6 and 7 show the values of λ and χ, the compressional components. A definite longitudinal compressional viscosity is indicated which however is somewhat smaller than the shear viscosity η.

V. MEASUREMENTS FOR SOLID POLYMERS

Recently two new methods have been devised for accurately measuring the properties of solid plastics in the ultrasonic frequency range

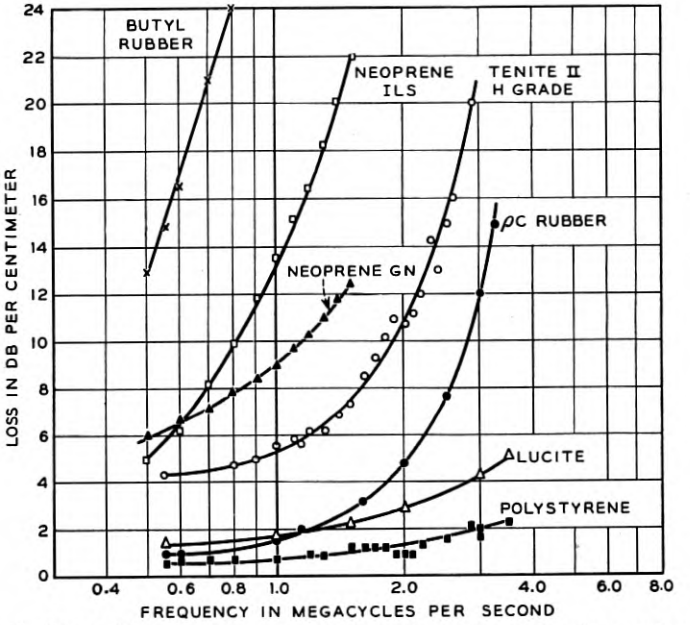


Fig. 24—Normal loss in db per centimeter measured as a function of frequency for several rubbers and plastics.

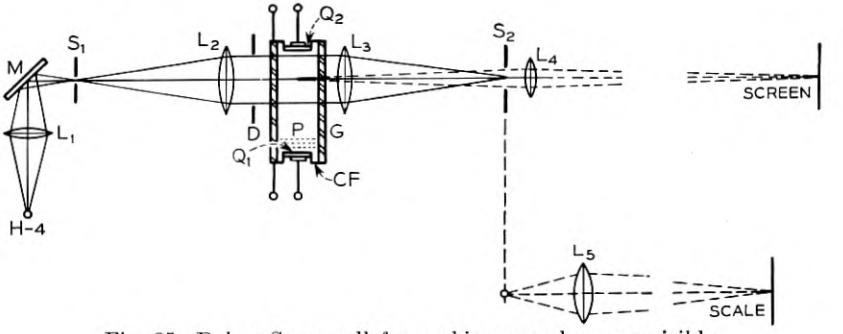


Fig. 25—Debye-Sears cell for making sound waves visible.

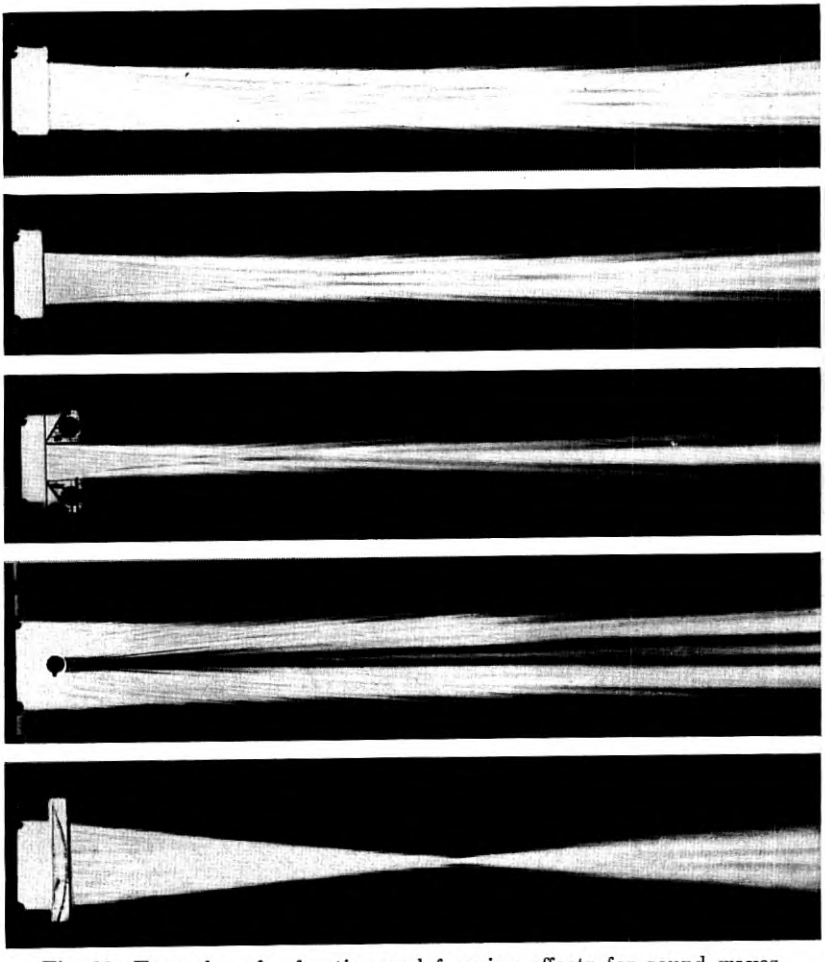


Fig. 26—Examples of refraction and focusing effects for sound waves.

and these have shown relaxations in such plastics as polyethylene and nylon 6-6. The simplest method for measuring one of the properties for longitudinal waves, i.e., the attenuation, is to measure the change in loss between two transducers in a liquid such as water, caused by inserting a sheet of the material. This process, used during the last war, results in the losses in db per centimeter for several rubbers and plastics, shown by Fig. 24. Indications of relaxation mechanisms are given by the rubbers and the plastic tenite II which is a cellulose acetate butyrate. The first fairly accurate method for measuring longitudinal sound

velocities[12] in plastics was the method of observing the focusing effect of a cylindrical lens made of the plastic. Sound waves can be made visible by the Debye-Sears technique of using a sound wave as a phase diffraction grating. Here light from a slit S_1 is made parallel by the lens L_2 and passes through the cell parallel to the wave fronts of the sound waves as shown by Fig. 25. The compressed parts of the medium retard the light waves more than the rarefied parts do and hence the medium acts as a phase diffraction grating. If a second slit S_2 is used which is small enough to pass only the zero order, a light valve action is obtained which modulates the light according to the sound wave intensity. If now the lens L_5 is used which focuses on the median plane of the tank, a picture of the sound beam is obtained as shown on Fig. 26. The bottom figure shows the focusing effect of a plastic and from the focal distance d and the radius of curvature r of the lense, one can calculate the velocity in a plastic compared to the velocity in the water by the formula

$$v_p = v_w \Big/ \left(1 - \frac{r}{d}\right) \tag{11}$$

This method gives velocities good to from 2 to 5 per cent depending on the attenuation in the lens.

G. W. Willard[13] has devised recently a more accurate method for measuring sound velocities as shown schematically by Fig. 27. Here a plastic to be measured is placed half way across the sound beam in the liquid and light is sent along the wave front occurring in both the plastic and the liquid. If the waves are in phase the retardation in the two light gratings, corresponding to sound propagation in both media, add up and for a slit selecting the zero order the darkest pattern occurs on the photographic plate. If the two waves are just out of phase, the retardation is reversed in the two media and the lightest part occurs. With this relation it can be shown[13] that the spacing d of light and dark lines

[12] W. P. Mason, *Piezoelectric Crystals and Their Application to Ultrasonics*, D. Van Nostrand, 1950, p. 404. It was used in this country by G. W. Willard as early as 1940. It was also used in Germany by J. Schaefer "Eine Neue Method Zur Messung der Ultraschallwellen in Festkorpern." Diss Strassburg, 1942. By making the front surface part of a cylinder, Schaefer also measured the shear velocity in a solid.

[13] G. W. Willard, *J. Acous. Soc. Am.*, **23**, Jan. 1951, pp. 83–94. The origin of this multiple path interference method goes back to the work of R. Bär (Helvetia Physica Physica Acta Bd 13 page 61 (1940)) who attached a piezoelectric crystal to a bar with a 45° end section and set up transverse and longitudinal waves, in the bar. These waves produced longitudinal waves in a surrounding liquid and by observing the interference pattern between them, the longitudinal and shear constants could be determined for an isotropic medium. Willard's method as described above is much more direct and is capable of higher accuracies.

is related to the wavelength in the liquid λ_ℓ and the wavelength in the solid λ_s, by

$$\frac{1}{d} = \frac{1}{\lambda_\ell} - \frac{1}{\lambda_s} \tag{12}$$

This corresponds to a velocity in the solid compared to a velocity in the liquid given by

$$v_s = \frac{v_\ell}{1 - v_\ell/fd} \tag{13}$$

where f is the frequency. Fig. 28 shows a photograph of a series of lines in a transparent plastic and a transparent plastic in the form of a wedge. It is seen that beyond the edge of the plastic there is a dark interference band for each one in the transparent plastic. This phenomenon is caused by the refraction of the sound wave that has traversed the plastic and the dark lines are lines of equal phase of the two waves in the liquid. The angle of the dark lines is half the refraction angle. Hence the velocity can also be determined by counting the number of dark bands in the liquid beyond the plastic. This makes it possible to measure the veloci-

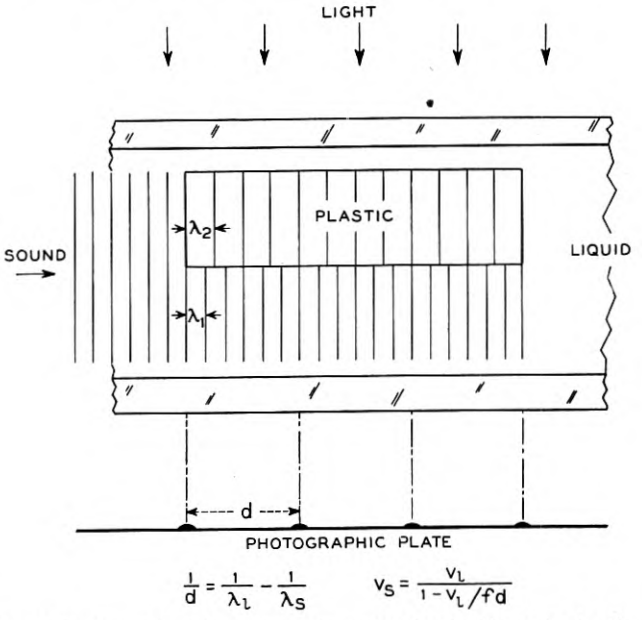


Fig. 27—Optical method for measuring sound velocities of plastics by comparing their velocity with that of a liquid such as water.

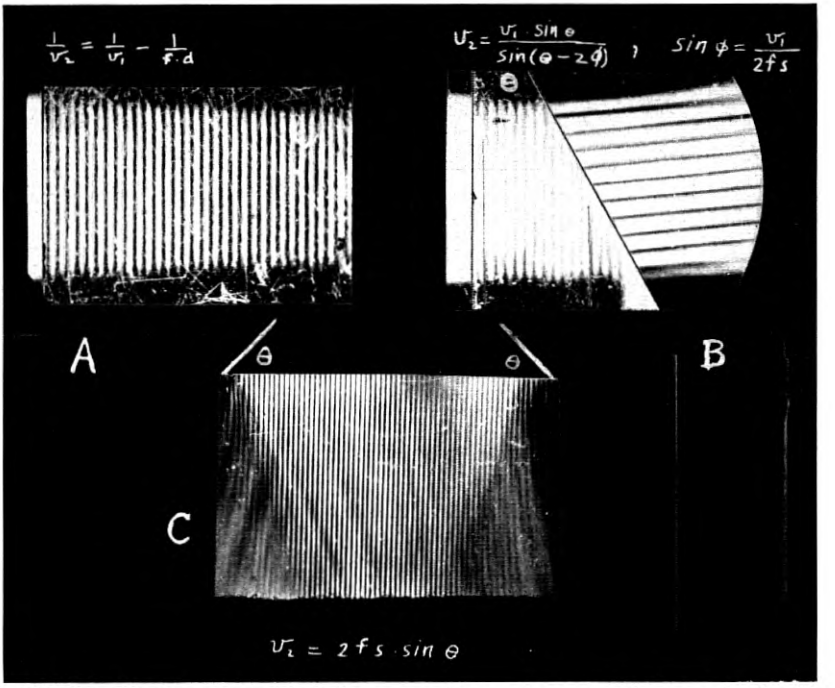


Fig. 28—Photographs of interference patterns from sound waves in liquids and plastics.

ties in opaque plastics. The accuracy of the method is better than 1 per cent if the attenuation is low enough to give a number of interference lines. For plastics of high internal loss, the method becomes somewhat inaccurate.

Typical measurements using this system are shown in Table V. Small changes in chemical composition and plasticizer content are shown up as can be seen from the table. Of particular interest is the difference between nylon 6-6 and polyethylene. Chemically as shown by Fig. 29, the two are identical except for the dipoles occurring for every 6 units of the ethylene chain. These dipoles have the effect of bonding adjacent layers together and result in a higher shearing modulus.

By attaching shear vibrating crystals to a right angled prism, as shown by the lower part of Fig. 28, with the direction of motion of the crystals parallel to the transmitting face, and setting up shear standing waves between the two crystals, the shear properties of the plastic can be measured. Longitudinal waves are generated in the liquid which interfere with one another and cause dark bands perpendicular to the plastic

TABLE V

Measured longitudinal velocity and attenuation at 25°C and 2.5 mc. Longitudinal attenuation in DB/cm at 25°C and 2.5 mc except as noted.

Material	Long velocity $\times 10^{-5}$ cm/sec	Shear velocity 10^{-5} cm/sec	A DB/cm	Density
Dural, 17 ST	6.5	3.12	—	2.7
Brass, half hard	4.7	2.11	—	—
Polystyrene	2.35	1.12	2	1.05
Plexiglas	2.68		5	1.18
Tenite II, (cellulose acetate butyrate), 2% plasticizer	2.08		9	1.23
Tenite II, 13% plasticizer	2.02		10	1.21
Polyvinal formal	2.68		10	1.24
Polyvinlidene chloride	2.4		18	1.71
Poly N-butyl methacrylate	1.96		5	1.05
Poly I-butyl methacrylate	2.08		6	1.05
Neoprene	1.51		20	.99
Polyethylene	2.0		4.7 $F^{1.11}$	.90
Nylon 6-6 (3–30 megacycles)	2.68		1.0 $F^{1.5}$	1.11
Nylon 6-10 (3–30 megacycles)	2.56		1.0 $F^{1.5}$	1.11

surface. By determining the spacing of these lines the velocity of the shear waves can be determined.

Another method has also been developed which is more applicable for high loss materials. This is a pulsing method and is a modification of the method proposed by one of the authors for measuring the properties of small crystal specimens.[14] Here longitudinal or shear crystals are soldered to the fused quartz rod as shown by Fig. 30 and a sample to be measured is placed between these by means of a liquid such as polyisobutylene which has a high shear elasticity. If the specimen has a small attenuation, this can be measured by taking the difference in the amplitude of successive reflections. If the specimen has a high loss, this does not work and another method has been used which consists in sending a pulse from both crystals.[15] One crystal is then used to receive and it receives the wave sent through the sample and the wave reflected from the fused quartz-sample interface. By adjusting the amplitude until these two are equal and the frequency or phase of one channel until the waves cancel, a ratio of amplitudes and a frequency of half wavelength are accurately determined. From these the velocity and attenuation can be calculated.

This method has been applied to measuring the longitudinal and shear velocities of polyethylene and 6-6 nylon. The polyethylene was of "equilibrium" crystallinity and average molecular weight corresponding to

[14] H. J. McSkimin, "Ultrasonic Measurement Techniques Applicable to Small Solid Specimens," *J. Acoust. Soc. Am.*, **22**, No. 4, July 1950, pp. 413–418.

[15] H. J. McSkimin, *J. Acoust. Soc. Am.*, **23**, No. 4, pp. 429–435.

an intrinsic viscosity in xylene of [n] = 0.89 at 85°C. Fig. 31 shows the longitudinal velocity of polyethylene plotted as a function of frequency and temperature. The velocity rises with frequency and a dispersion is indicated. This is confirmed by the attenuation per wavelength curve for two different frequencies plotted as a function of temperature, Fig. 32. A definite dispersion is seen to occur with an activation energy of about 12 ± 2 kilocalories per mole. This could occur in either the λ constant or the shearing constants μ, but the data of Figs. 33 and 34 show definitely that it occurs in the shear component. Fig. 33 shows the shear velocity for four temperatures plotted as a function of frequency. This can be fitted for 30°C with a single relaxation mechanism having a relaxation frequency of 8 megacycles. To agree with the measured attenuation and velocity, there has to be a spreading of the single relaxation over a range as also occurs in liquids. The indicated shear stiffness below this relaxation frequency is 2.6×10^9 dynes/cm^2. Some

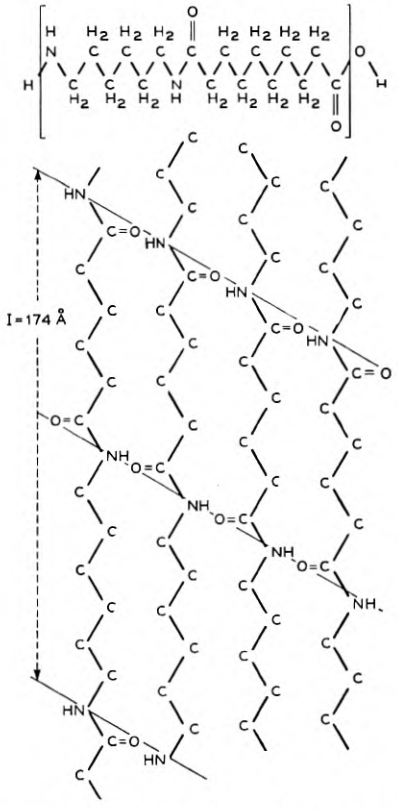


Fig. 29—Spatial structure of nylon 6-6.

data on the zero frequency shear modulus is obtained from the Young's modulus for a static pull which is from 30,000 to 50,000 pounds/square inch. Since the Young's modulus is three times the shearing modulus, the zero frequency shearing modulus should not exceed 1.1×10^9 dynes/cm^2. Hence one may expect that other relaxations will occur at lower frequencies.

Fig. 34 shows the attenuation per wavelength for shear waves. The solid line for 30°C represents the calculated attenuation per wavelength for the model assumed. If all the dissipation were due to shear mechanisms, the calculated attenuations would occur as shown by the 30°C

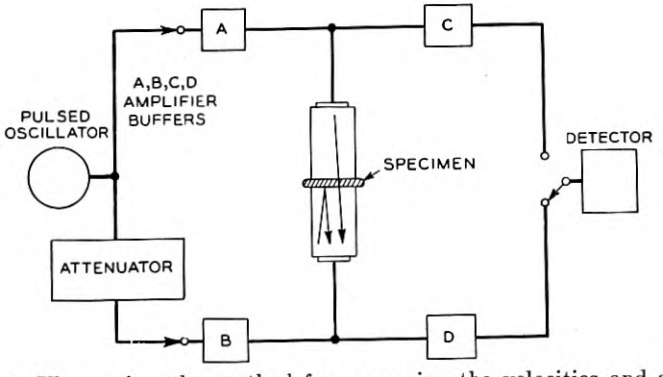


Fig. 30—Ultrasonic pulse method for measuring the velocities and attenuations of highly attenuating plastics.

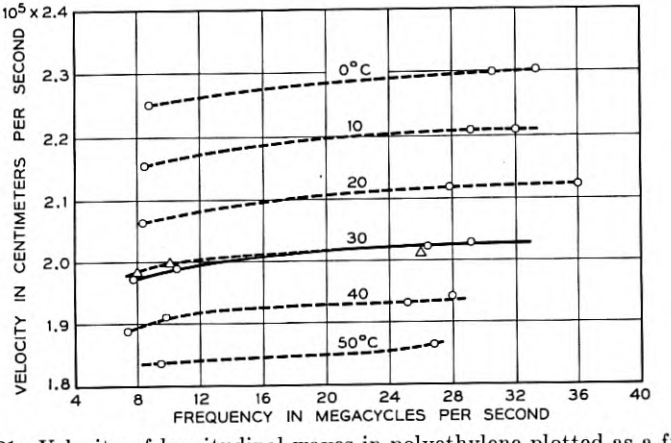


Fig. 31—Velocity of longitudinal waves in polyethylene plotted as a function of temperature and frequency.

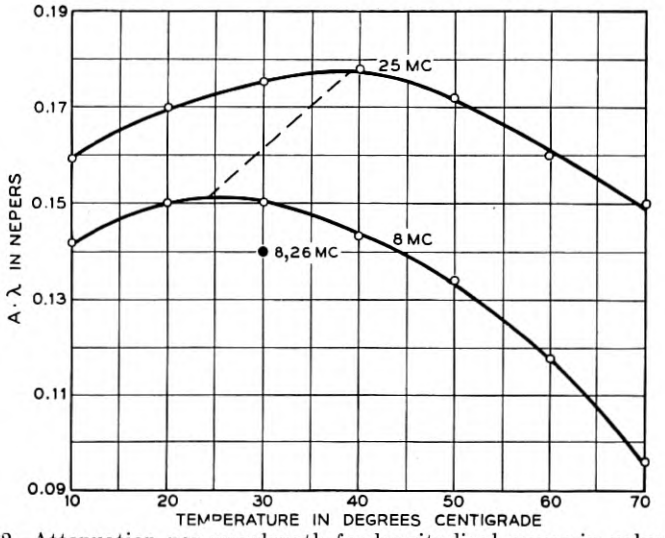


Fig. 32—Attenuation per wavelength for longitudinal waves in polyethylene plotted as a function of temperature and frequency.

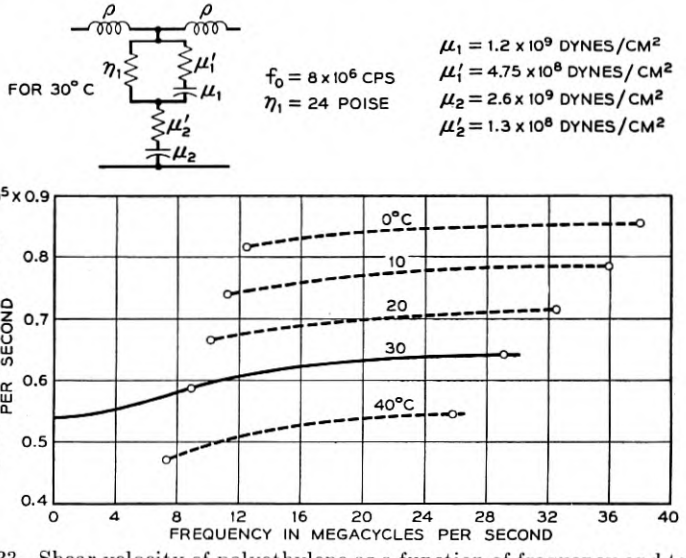


Fig. 33—Shear velocity of polyethylene as a function of frequency and temperature. Equivalent circuit shows elements necessary to account for the velocity and attenuation changes at 30°C.

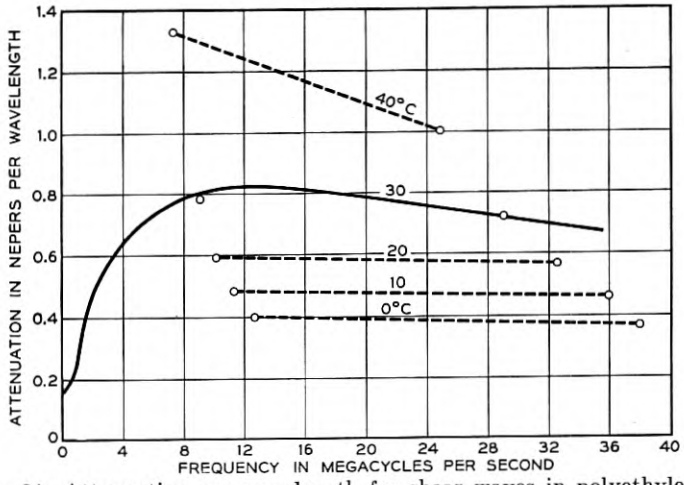


Fig. 34—Attenuation per wavelength for shear waves in polyethylene.

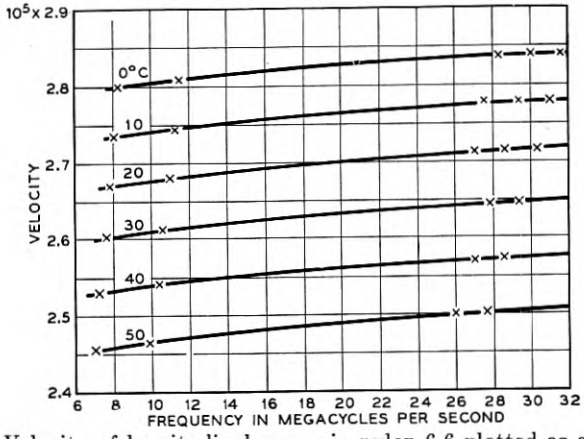


Fig. 35—Velocity of longitudinal waves in nylon 6-6 plotted as a function of temperature and frequency.

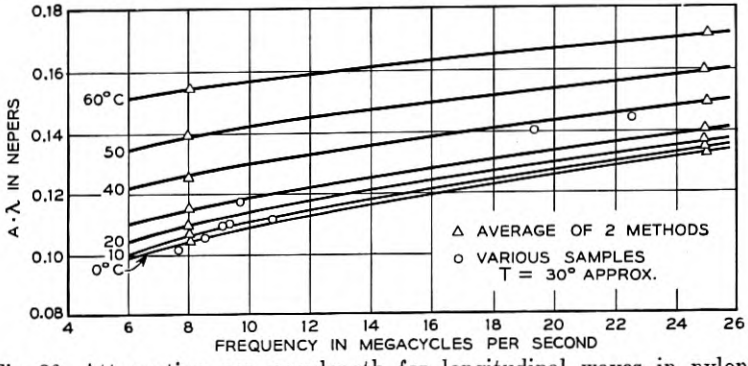


Fig. 36—Attenuation per wavelength for longitudinal waves in nylon 6-6 plotted as a function of temperature and frequency.

points of Fig. 32. Most of the loss is accounted for by shear mechanisms, but it appears that some compressional mechanisms may also be present.

The mechanism causing the relaxation in the megacycle range for polyethylene appears to be the same as for polyisobutylene, namely the relaxation of the shortest chain segment that is free to move. The chain segment acting appears to be longer than six chain units for similar measurements of nylon 6-6 show no relaxations in this frequency range. Fig. 35 shows the longitudinal velocity and Fig. 36 the attenuation per wavelength for longitudinal waves. Since the attenuation per wavelength is still increasing for nylon 6-6 at 25 megacycles a still shorter chain segment may be operating for this material. The shear velocity and attenuation per wavelength for nylon 6-6 are shown by Figs. 37 and 38.

Fig. 39 shows the shear stiffness of polyethylene and nylon 6-6 plotted

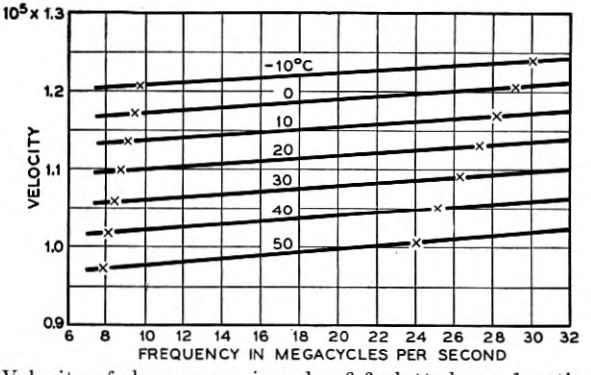


Fig. 37—Velocity of shear waves in nylon 6-6 plotted as a function of temperature and frequency.

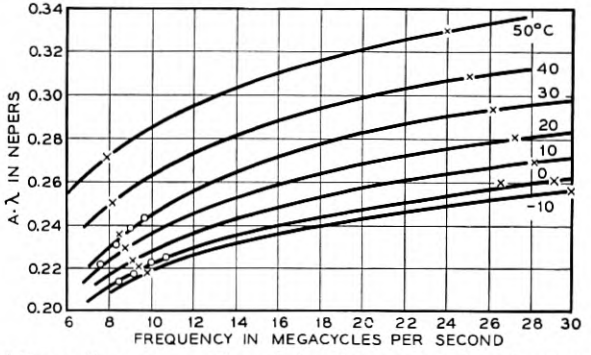


Fig. 38—Attenuation per wavelength for shear waves in nylon 6-6 plotted as a function of temperature and frequency.

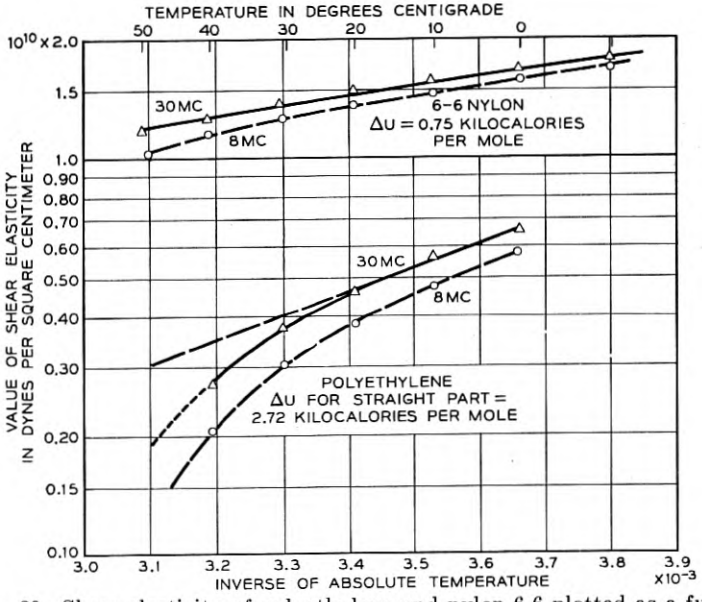


Fig. 39—Shear elasticity of polyethylene and nylon 6-6 plotted as a function of temperature and frequency.

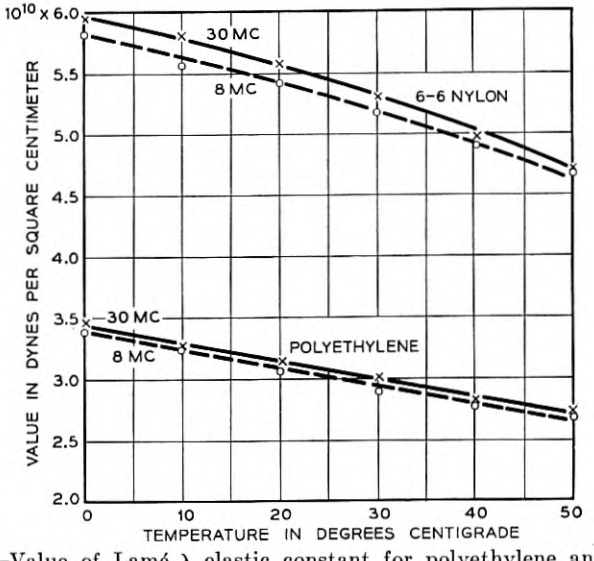


Fig. 40—Value of Lamé λ elastic constant for polyethylene and nylon 6-6 plotted as a function of frequency and temperature.

against $1/T$ where T is the absolute temperature. Both are plotted for 8 mc and 30 mc. The dispersion in both materials is evident. Below 30°C the shear elasticity of polyethylene varies exponentially with the temperature with an activation energy of 2.72 kilocalories per mole. Above this temperature a deviation occurs due to the approach to the melting temperature. Nylon has a smaller variation with temperature.

Comparing the longitudinal and shear wave measurements one can calculate the Lamé λ elastic constant and this is shown plotted on Fig. 40 for both polyethylene and nylon 6-6 as a function of temperature for

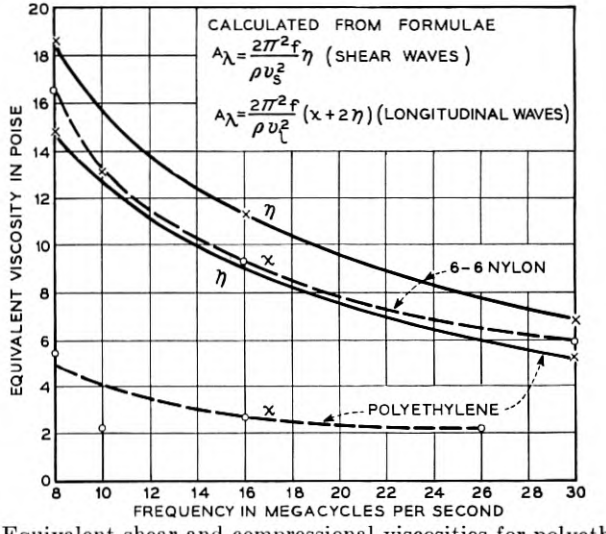


Fig. 41—Equivalent shear and compressional viscosities for polyethylene and nylon 6-6 plotted as a function of frequency for a temperature of 25°C.

two frequencies. The dispersion of λ for polyethylene is small but is more prominent in nylon 6-6. This correlates with the larger compressional viscosity component present for nylon 6-6 which as shown from Fig. 41 is as large as the shear viscosity. According to the structural rearrangement theory of compressional viscosity due to Debye,[16] compressional viscosity can enter when some part of the chain can rearrange from one stable state to another stable state as a function of pressure. This rearrangement occurs across a potential barrier and hence requires a finite amount of time to occur. This lag in the rearrangement results in a compressional viscosity and as the frequency is increased, a frequency is found for which the motion can no longer occur in the time of a single

[16] P. Debye, *Z. Elektrochem.*, **45**, 1939, p. 174.

cycle and the λ constant increases. It appears from these measurements that the dipole binding present in nylon 6-6 allows a greater structural rearrangement under pressure than can occur for polyethylene which has only linear chains.

VI. CONCLUSIONS

Measurements in dilute solutions, in pure polymer liquids and in non rigid solid polymers have all shown the presence of a shortest segment whose relaxation leads to a crystalline type of elasticity. In dilute polymer solutions the presence of a configurational type of relaxation and an entanglement relaxation of the shortest chain segment have been shown. For pure polymer liquids a quasi-configurational type of relaxation has been found for chain lengths greater than 60 segments, but for chain lengths less than 40 segments this type of relaxation disappears. From the difference between the high frequency shear elasticities measured for polyethylene and nylon 6-6 and the static measurement of Young's modulus, it appears that there may be other relaxations in these materials for lower frequency ranges.

For pure polyisobutylene and for nylon 6-6 there appear to be structural changes induced by pressure which account for a compressional viscosity and a dispersion in the λ elastic constant. This effect is smaller for polyethylene.

APPENDIX—EFFECT OF LIQUIDS ON THE PROPAGATION OF SHEAR WAVES IN RODS

For radially symmetric rods, the tangential particle displacement u_θ in the rod is given by

$$u_\theta = J_1(kr)e^{j\omega t-\theta z} \tag{1}$$

where

$$k^2 = \frac{\rho\omega^2}{\mu} + \theta^2 \tag{1A}$$

All other displacements are zero. In this equation waves are considered to be travelling in the $+z$ direction with a propagation constant $\theta = A + jB$, where A is the attenuation in nepers per cm and B the phase shift in radians per centimeter. μ is the shear stiffness which may be complex to take account of the dissipation within the rod.

From the defining relations for the stress strain equation

$$T_{r\theta} = \mu S_{r\theta} = \mu \left(\frac{\partial u_\theta}{\partial r} - \frac{u_\theta}{r} \right)$$

and the tangential particle velocity $\partial u_\theta / \partial t$, one can calculate the impedance Z per square cm. of cylindrical surface at $r = a$. This relation is

$$Z = \frac{-T_{r\theta}}{\dot{u}_\theta} = \frac{j\mu k}{\omega} \left[\frac{J_0(ka)}{J_1(ka)} - \frac{2}{ka} \right] \tag{2}$$

Since only the first mode is excited, parameters can be adjusted to keep k quite small, i.e. ($ka < .2$) and equation (2) can be simplified by using power series expansions for the Bessel functions. Neglecting higher order terms this results in

$$Z = \frac{-j\mu a k^2}{4\omega} \tag{3}$$

To evaluate the impedance of the liquid surrounding the rod, the torsional wave is first propagated along the length of the rod without the liquid, i.e. with $Z = 0$. Then from equation (3) $k = 0$ and from equation (1A)

$$\theta_0^2 = \frac{-\rho\omega^2}{\mu} = (A_0 + jB_0)^2 \tag{4}$$

where A_0 and B_0 are respectively the attenuation and phase shift in the rod alone. With the small loss in metal and glass rods A_0 can be taken equal to zero and

$$B_0 = \omega / \sqrt{\mu/\rho} = \omega / v_0 \tag{5}$$

where v_0 is the velocity of propagation in the rod alone.

When the liquid surrounds the rod, however,

$$k^2 = \frac{\rho\omega^2}{\mu} + \theta^2 = -(A_0 + jB_0)^2 + (A + jB)^2 \tag{6}$$

For the usual case where $(B + B_0) \gg (A + A_0)$, equation (6) approximates

$$k^2 = (B + B_0)(-\Delta B + j\Delta A) \doteq 2B_0(-\Delta B + j\Delta A)$$

where ΔB is the increase in phase shift per centimeter and ΔA the increase in attenuation per cm, both directly measurable quantities. The

final working equation is then given by

$$Z = \frac{\mu a}{4\omega} \times 2B_0[\Delta A + j\Delta B] = \frac{\mu a}{2v_0}(\Delta A + j\Delta B) = \frac{\rho v_0 a}{2}(\Delta A + j\Delta B) \quad (7)$$

Since ΔA and ΔB are the attenuation and phase shift changes per unit length, then if ℓ is the length of the rod, covered the total attenuation and phase shift changes will be ΔA and ΔB multiplied by 2ℓ. Hence if $\overline{\Delta A}_0$ and $\overline{\Delta B}_0$ are the measured attenuation and phase changes, the impedance Z becomes

$$Z = \frac{\rho v_0 a}{4\ell}(\overline{\Delta A}_0 + j\overline{\Delta B}_0) \quad (8)$$

This derivation neglects the change of phase occurring at the intersection between the rod having no liquid and the rod surrounded by the liquid, but it can be shown that this is small and moreover, the change in the wave on leaving is equal and opposite to that occurring on entering and hence this correction cancels out. However, if the liquid is viscous enough there is a correction due to the fact that the measured impedance of equation (8) is for a cylindrical surface, whereas the desired impedance is the characteristic plane wave impedance. Obviously if the radius of curvature is sufficiently large no correction to equation (8) need be made. To obtain a suitable criterion one may consider waves propagated into the liquid from the surface of the rod and solve for the impedance per square cm. of the cylindrical surface. This neglects the variation with z, but since the wavelength along the rod is quite large, little error results from neglecting variations with z.

An outgoing cylindrical wave in the medium may be represented by

$$u_\theta = [J_1(kr)' - jY_1(kr)']e^{j\omega t} = H_1^{(2)}(kr)'\, e^{j\omega t} \quad (9)$$

where the primes refer to the wave in the liquid and

$$(k')^2 = \frac{\rho'\omega^2}{\mu'} = \frac{\rho'^2\omega^2}{Z_k^2} \qquad \text{or} \qquad k'a = \frac{\rho'\omega a}{Z_k} \quad (10)$$

where $Z_k = \sqrt{\mu'\rho'}$ is the plane wave impedance of the liquid.

The shearing stress

$$T_{r\theta} = \mu' S_{r\theta} = \mu'\left[\frac{\partial u_\theta}{\partial r} - \frac{u_\theta}{r}\right],$$

and the tangential velocity may be obtained as before and the complex impedance over the cylindrical surface determined to be

$$Z = \frac{-T_{r\theta}}{\dot{u}_j} = -\mu' \left[\frac{\partial H_1^{(2)}(ka)'}{\partial r} - \frac{H_1^{(2)}(kr)'}{r} \right] \Big/ j\omega H_1^{(2)}(kr)' \quad (11)$$

Noting that for plane waves $Z_k = \sqrt{\rho'\mu'}$, one may eliminate μ' from (11) and evaluate Z at $r = a$ with the result

$$Z = j \left[\frac{H_0^{(2)}(ka)'}{H_1^{(2)}(ka)'} - \frac{2}{(ka)'} \right] Z_k \quad (12)$$

The above equation can be used to obtain a solution for Z_k in terms of the measured value of the cylindrical impedance Z for any set of parameters that may apply. Except for very heavy loading of the rod, however, the results of equation (8) may be used directly with little error. For example calculations indicate that for $| ka |' = 10$, the multiplier of Z_k of equation (12) is approximately $1.07 \angle -7.4°$ for phase angle of $(ka)'$ of $-25°$. For $(ka)' < 10$, the correction multiplier rapidly becomes important. This same correction is applicable for the torsional crystal, but since this is only used for dynamic viscosities less than 10 poises, a correction is seldom necessary.

Relay Armature Rebound Analysis

BY ERIC EDEN SUMNER

(Manuscript received October 25, 1951)

Rebound of mechanical structures subsequent to impinging on stops generally has deleterious effects on their performance and should, therefore, be minimized. A considerable reduction in rebound can often be obtained by introducing additional degrees of freedom to the structure.

A mathematical treatise of the dynamics of rebound motion of systems representing idealized relay armatures is presented. Normalized differential equations of motion and their solutions for the "free" and "impact" intervals are derived for systems having one, two, and three degrees of freedom, allowing the rebound behavior of a specific system to be calculated. The equations of series of rebounds, and possible combinations of such series are considered next for systems having one and two degrees of freedom. The field of possible rebound maxima is mapped for a practical range of mass distribution constants, coefficients of restitution, and force ratios. A sufficiently broad optimum design region is indicated.

The results of this analysis have been checked closely on a model and have led to appreciable reduction of armature rebound in relay designs.

I. INTRODUCTION

In numerous types of mechanisms it is desirable to arrest the motion of a member at a particular point and to maintain it in this position. One of the simplest means of accomplishing this is to allow the moving member to impinge on a fixed member (stop) and to provide forces to tension it against this stop. Because the member to be arrested possesses kinetic energy and because the stop cannot generally absorb all of this energy, the moving member will rebound from the stop. The rebound motion generally deteriorates the performance of the mechanism and should be minimized.

Investigation of this phenomenon has been stimulated by the armature rebound problem in relay operation, where rebound from the front stop* tends to reclose contacts and must therefore be compensated for by additional (waste) travel, resulting in deleterious effects on speed and

* Among relay designers the front stop has been generally referred to as "backstop". In this paper the terms front stop and heel stop have been used throughout for easier identification.

magnetic characteristics. Analysis in this paper will be directed towards relay armature systems, but it is also applicable to rebound in similar mechanisms.

II. STATEMENT OF PROBLEM

Analysis will be restricted to planar motion of armature systems having one, two, and three degrees of freedom as depicted in Figs. 1, 2, and 3. Generally one stop must be provided for each degree of freedom, although in the three-degree-of-freedom system of Fig. 3, two of the stops have been combined.

Applied forces F_1, F_2, F_3, have been chosen so as to be most easily correlated with actual relay designs.

The initial condition in all cases will be a pure rotation about the "heel" just prior to a "zero" impact at the "front" of the armature. The "zero" impact will be followed by rebound motion and impacts at the various stops eventually bringing the armature to rest. The object

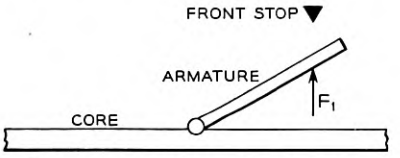


Fig. 1—Solidly hinged armature—one degree of freedom.

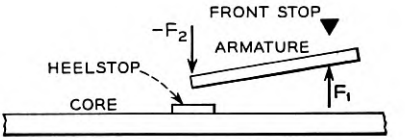
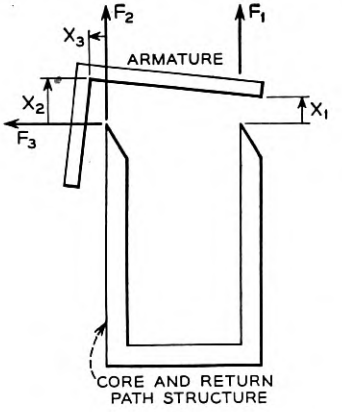
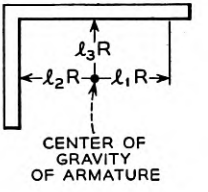


Fig. 2—Loosely hinged armature—two degrees of freedom.



(a)



(b)

Fig. 3—Armature system—three degrees of freedom.

will be to minimize rebound motion at the front, since this is usually near the point actuating the relay contacts.

The basic problem is then to find the response of the armature subject to aperiodic but well defined impulses, which are functions of the positions and velocities of the system.

III. ASSUMPTIONS

In order to facilitate the solution of this problem, the following modifying assumptions are made:

(1) As mentioned in the previous section, analysis is restricted to planar motion.

(2) The armature is assumed to be a rigid body.

(3) Stops are assumed to be very stiff, massless springs capable of energy absorption during impact with the armature. The associated coefficient of restitution is assumed constant. Core and stop vibration are neglected.

(4) The tensioning forces F_1 , F_2 , F_3 are assumed to be constant forces. (This is fairly closely true for moderate rebound amplitudes of practical relay structures.)

(5) All displacements are small relative to the dimensions of the system and in particular the angular displacement θ is sufficiently small so that

$$\cos \theta \doteq 1$$

$$\sin \theta \doteq \theta$$

IV. DERIVATION OF EQUATIONS OF MOTION

The derivation of the equations of motion resolves itself into the solution of two different types of intervals:

(1) *Free Interval:* This is the period during which the armature is not in contact with any of its stops and only the tensioning forces are acting.

(2) *Impact Interval:* During such intervals the armature is in contact with at least one of the stops. The stiffness of the latter is assumed so high that the tensioning forces during this interval may be neglected.

The three-degree-of-freedom case will be considered first and the others subsequently deduced from it by allowing some of the constants to approach zero.

A. Free Interval

The motion of the armature will be described by the displacement at the stop points: x_1, x_3, x_3.* Let m be the mass and R the radius of gyration of the armature about the center of gravity. The latter is located by the dimensions $\ell_1 R$, $\ell_2 R$, and $\ell_3 R$ relative to the stop points, i.e., the points on the armature which contact the stops in the rest position (Fig. 3).

The equations of motion are derived in Appendix I and are put into dimensionless form:

$$\left.\begin{aligned} y_1 &= \frac{1}{2}\left[C_{11} + C_{12}\frac{(F_2)}{(F_1)} + C_{13}\frac{(F_3)}{(F_1)}\right]\left(\frac{t}{\tau}\right)^2 + \dot{y}_{10}\left(\frac{t}{\tau}\right) + \dot{y}_{10} \\ y_2 &= \frac{1}{2}\left[C_{21} + C_{22}\frac{(F_2)}{(F_1)} + C_{23}\frac{(F_3)}{(F_1)}\right]\left(\frac{t}{\tau}\right)^2 + \dot{y}_{20}\left(\frac{t}{\tau}\right) + \dot{y}_{20} \\ y_3 &= \frac{1}{2}\left[C_{31} + C_{32}\frac{(F_2)}{(F_1)} + C_{33}\frac{(F_3)}{(F_1)}\right]\left(\frac{t}{\tau}\right)^2 + \dot{y}_{30}\left(\frac{t}{\tau}\right) + \dot{y}_{30} \end{aligned}\right\} \quad (1)$$

where:

$$\left.\begin{aligned} y_i = \frac{x_i}{\dot{x}_a \tau} = \frac{F_1}{\dot{x}_a^2 m}x_i \qquad \dot{y}_i = \frac{d}{d\left(\frac{t}{\tau}\right)}y_i = \frac{\dot{x}_i}{\dot{x}_a} \\ \tau = \frac{\dot{x}_a m}{F_1} \end{aligned}\right\} \quad (2)$$

$\dot{x}_a$ is the front velocity $\dot{x}_1$, just prior to the "zero" impact, and

$$\left.\begin{aligned} C_{11} &= (\ell_1^2 + 1) \quad C_{13} = C_{31} = \ell_1\ell_3 \\ C_{22} &= (\ell_2^2 + 1) \quad C_{12} = C_{21} = (1 - \ell_1\ell_2) \\ C_{23} &= (\ell_3^2 + 1) \quad C_{23} = C_{32} = -\ell_2\ell_3 \end{aligned}\right\} \quad (3)$$

$\dot{y}_{10}$, $\dot{y}_{20}$, $\dot{y}_{30}$, are the initial velocities and y_{10}, y_{20}, y_{30} the initial displacements for the free interval in question.

The equations of motion for a two-degree-of-freedom system are obtained, if $F_3 = 0$. Then for the two coordinates of interest:

$$\left.\begin{aligned} y_1 &= \frac{1}{2}\left[C_{11} + C_{12}\left(\frac{F_2}{F_1}\right)\right]\left(\frac{t}{\tau}\right)^2 + \dot{y}_{10}\left(\frac{t}{\tau}\right) + y_{10} \\ y_2 &= \frac{1}{2}\left[C_{21} + C_{22}\left(\frac{F_2}{F_1}\right)\right]\left(\frac{t}{\tau}\right)^2 + \dot{y}_{20}\left(\frac{t}{\tau}\right) + y_{20} \end{aligned}\right\} \quad (4)$$

* A summary of all notations used in this paper is given in Appendix IV.

For a one-degree-of-freedom system $\ddot{y}_2 = C_{21} + C_{22}\left(\frac{F_2}{F_1}\right) = 0$, whence

$$y_1 = \frac{1}{2}\left[C_{11} - \frac{C_{12}^2}{C_{22}}\right]\left(\frac{t}{\tau}\right)^2 + \dot{y}_{10}\left(\frac{t}{\tau}\right) + y_{10}\Bigg\} \tag{5}$$

B. Impact Interval

The change of velocity at point "i" due to an impact at "i" is, by definition of the coefficient of restitution "k",

$$\Delta\dot{x}_1 = -(1 + k_i)\dot{x}_i$$

It is assumed here that the action of the stops are true impacts, i.e., the changes in velocity take place while there is negligible motion of the body. The velocity changes then occur as instantaneous rotation about the conjugate axis, leading to the general relation for an impact at point "i":

$$\dot{y}_{j0n} = \dot{y}_{je(n-1)} + K_{ji}\dot{y}_{ie(n-1)} \tag{6}$$

The first subscript indicates the coordinate, the second subscript indicates the beginning (0) or the end (e) of the free interval described by the third subscript. The impact transfer coefficient K_{ji} relating a velocity change at point "j" to an impact at point "i":

$$K_{ji} = -\frac{C_{ji}}{C_{ii}}(1 + k_i) \tag{7}$$

Equations (1) through (7) allow any one specific case to be mapped, if the mass distribution and force ratio are known. A sample of such mapping of rebound motion for a rectangular two-degree-of-freedom armature appears in Fig. 4.

V. ANALYSIS OF REBOUND PATTERN—ONE-DEGREE-OF-FREEDOM SYSTEM

The rebound pattern for the one-degree-of-freedom system—as derived in Appendix II—consists of an infinite series of parabolic arcs of diminishing amplitudes. The structure comes to rest after a finite time interval. The maximum rebound occurs during the first bounce and equals

$$Y = -\frac{k^2}{2C} \tag{8}$$

where

$$C = C_{11} - \frac{C_{12}^2}{C_{22}} \tag{9}$$

The system returns to rest at

$$\frac{t}{\tau} = \frac{2}{C(1 - k)} \tag{10}$$

VI. ANALYSIS OF REBOUND PATTERN—TWO-DEGREE-OF-FREEDOM SYSTEM

The reason for choosing a two-degree-of-freedom system over a one-degree-of-freedom system would be, in keeping with the philosophy of

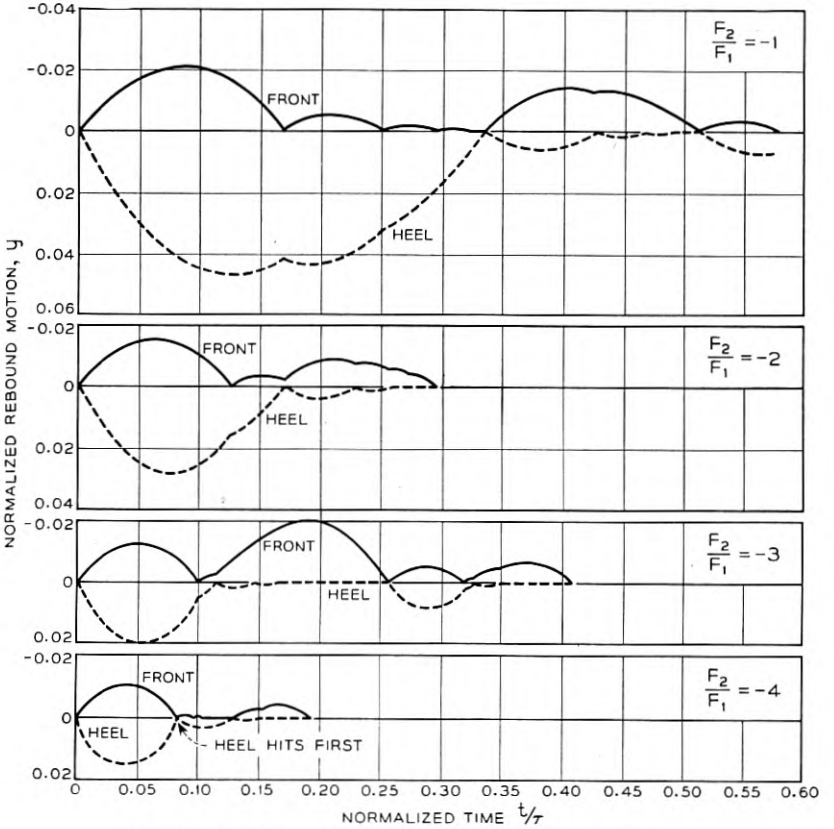


Fig. 4—Front and heel motion of plate type armature.

this treatment, to reduce Y_1, the greatest excursion at the front. In order to simplify mapping, this maximum excursion will be expressed as $2CY_1$, the ratio of Y_1 to Y as given by Equation (8) for the case of $k = 1$. Thus $2CY_1$ is the ratio of the greatest excursion of the two-degree-of-freedom system under consideration to the greatest excursion of the corresponding perfectly elastic one-degree-of freedom system.

We first introduce two basic constants which are functions of the mass distribution relative to the stop locations:

$$M_{ij} = \frac{C_{ij}^2}{C_{ii}C_{jj}} \tag{11}$$

This constant represents a mechanical coupling coefficient. As $M_{ij} = M_{ji}$, the two-degree-of-freedom system under consideration here has only one such non-trivial constant M_{12}.

The second constant represents a force transformation factor from the "j" coordinate to the "i" coordinate:

$$P_{ij} = \frac{C_{ij}}{C_{ii}} \tag{12}$$

In the analysis of the two-degree-of-freedom system only P_{12} is important.

If there is to be any heel motion, the "zero" impact at the front must impart a positive velocity to the heel. By Equations (6), (7), and (12), this requires that P_{12} be negative, which in turn implies that $\ell_1\ell_2 > 1$. For the limiting case of $\ell_1\ell_2 = 1$, $P_{12} = M_{12} = 0$ and no coupling exists between the heel and the front. Physically this means that the two stops are the centers of percussion of each other and the system will act as a simple hinge.

With the above foundation, it is possible to analyze the patterns of motion and maximum rebound amplitudes.

A. Motion Immediately Following "Zero" Impact

After the "zero" impact at the front, both front and heel will lift off in accordance with impact Equation (6) and continue to move in accordance with the free interval Equations (4). Whether the next impact occurs at the front or the heel depends on their respective periods, t_1 and t_2:

$$\frac{t_1}{t_2} = \frac{1 + \dfrac{P_{12}}{M_{12}} f}{1 + P_{12} f} \frac{k_1}{1 + k_1} \tag{13}$$

where:

$$f = \frac{F_2}{F_1}$$

A large value of t_1/t_2 will result in a series of heel impacts and the heel will come to rest while the front is still displaced from the stop. This will be called a complete heel series. A small value of t_1/t_2 results in a similar complete front series. If t_1/t_2 is near unity, a limited number of impacts on one end are followed by an impact on the other end, etc. An analysis of front and heel series follows:

B. Front Series

If $t_1/t_2 < 1$ a series of front impacts occurs. The impact velocities at the front are

$$\dot{y}_{1n} = 1, k, k^2, \cdots, k^n \tag{14}$$

The corresponding time intervals are

$$T_{1n} = \frac{2k_1}{A}, \frac{2k_1^2}{A}, \cdots, \frac{2k^n}{A} \tag{15}$$

where

$$A = (C_{11} + C_{12}f)$$

During this time, the heel velocity and displacement are given by

$$\left.\begin{aligned} \dot{y}_{20(n+1)} &= \dot{y}_{20n} + \left[\frac{2B}{A} - P_{12}(1 + k_1)\right] y_{i0n} \\ y_{20(n+1)} &= y_{20n} + \left[\frac{2B}{A}\dot{y}_{1n} + \dot{y}_{20n}\right]\dot{y}_{10n} \end{aligned}\right\} \tag{16}$$

where

$$B = C_{12} + C_{22}f$$

The velocity and displacement at the heel after a given number of front impacts are obtained by a summation of Equations (16). For a complete front series $n \to \infty$, and

$$\left.\begin{aligned} y_{2e\infty} &= \frac{2k_1}{A(1 + k_1)^2}\left[\frac{Bk_1}{A} - P_{12}\right] \\ \dot{y}_{2e\infty} &= \frac{1}{1 - k_1}\left[\frac{2Bk_1}{A} - P_{12}(1 + k_1)\right] \end{aligned}\right\} \tag{17}$$

In addition it is useful to set down energy equations in order to simplify evaluation of greatest rebound for the various groups of rebound patterns. The kinetic energy function T is evaluated in Appendix I. A potential energy term V—the work done against F_1 and F_2 from the equilibrium position—is introduced. If T_0 is the total energy of the system prior to the "zero" impact, then

$$\frac{T+V}{T_0} = \dot{y}_1^2 + \frac{M_{12}}{P_{12}^2}\dot{y}_2^2 - \frac{2M_{12}}{P_{12}}\dot{y}_1\dot{y}_2 - 2C(y_1 + fy_2) \tag{18}$$

The energy loss due to n front impacts is

$$-\Delta\left(\frac{T+V}{T_0}\right) = (1 - k_1^{2n})(1 - M_{12})\dot{y}_{1e0}^2 \tag{19}$$

For a complete front series $n \to \infty$, and

$$-\Delta\left(\frac{T+V}{T_0}\right) = (1 - M_{12})\dot{y}_{1e0}^2 \tag{20}$$

If a complete front series follows the "zero" impact, $\dot{y}_{1e0} = 1$ and

$$-\Delta\left(\frac{T+V}{T_0}\right) = (1 - M_{12}) \tag{21}$$

After completion of this "initial" front series, the system maintains only one degree of freedom (rotation about the front) until a heel impact occurs. By setting $\dot{y}_1 = y_1 = y_2 = 0$ in (21) we obtain the heel impact approach velocity $\dot{y}_2 = P_{12}$.

Apparently energy loss due to n front impacts is a function of M_{12}, k_1, and the approach velocity of the first impact.

C. Heel Series

An analysis similar to the above can be made for partial and complete heel series following the "zero" impact. This is demonstrated in Appendix III, yielding, for $k_1 = k_2$*

$$\left.\begin{aligned} y_{1e\infty} &= \frac{AP_{12}(1+k)^2}{B(1-k)^2}\left[\frac{AP_{12}}{B} - \frac{k(1-k)}{1+k} - M_{12}k\right] \\ \dot{y}_{1e\infty} &= \frac{1+k}{1-k}\left[\frac{2AP_{12}}{B} - \frac{k(1-k)}{(1+k)} - M_{12}(1+k)\right] \end{aligned}\right\} \tag{22}$$

* The more general form $k_1 \neq k_2$ can be obtained as indicated in Appendix III.

The energy relationships for heel series are

$$-\Delta\left(\frac{T+V}{T_0}\right) = (1 - k_2^{2n})\frac{M_{12}(1 - M_{12})}{P_{12}^2}\dot{y}_{2e0}^2 \tag{23}$$

For a complete series $n \to \infty$, and

$$-\Delta\left(\frac{T+V}{T_0}\right) = \frac{M_{12}(1 - M_{12})}{P_{12}^2}\dot{y}_{2e0}^2 \tag{24}$$

If a complete heel series follows the "zero" impact, $\dot{y}_{2e0} = P_{12}(1 + k_1)$, and

$$-\Delta\left(\frac{T+V}{T_0}\right) = M_{12}(1 - M_{12})(1 + k_1)^2 \tag{25}$$

Finally, for the special case where a complete heel series follows an initial complete front series $\dot{y}_{2e0} = P_{12}$, and

$$-\Delta\left(\frac{T+V}{T_0}\right) = M_{12}(M_{12} - 1) \tag{26}$$

It is to be noted that the energy loss due to a partial heel series is a function of M_{12}, P_{12}, k_2, and the approach velocity of the first impact, but that the equation for a complete heel series does not contain k_2. Finally, a complete initial heel series is a function of only M_{12} and k_1.

D. Complete Mapping of Problem

Equations (1) through (26) make it possible to completely map the two-degree-of-freedom rebound problem. The relative maximum amplitude $2CY_1$ and the rebound pattern will be determined.

Examination of the necessary equations, show that $2CY_1$ is in all cases a function of four parameters: k_1, k_2, M_{12} and $P_{12}f$. Of these, k_2 enters only if a partial heel series occurs prior to the time of maximum rebound. If it is assumed that for this limited group of cases $k_2 = k_1 = k$, the number of parameters is reduced to three: k, M_{12}, $P_{12}f$.

In Figs. 5 to 10, $2CY_1$ is plotted against $P_{12}f$ for the most useful range of $1/8 < M_{12} < 1/2.5$, $0.3 < k < 0.6$ and $0 < P_{12}f < 10$.

As $P_{12}f$ is increased from zero to infinity (corresponding to an increase in the heel tension F_2), the rebound pattern goes through some or all of five regions. The criterion for location in any one region is based upon the parameter

$$Q = \frac{1 + \frac{P_{12}}{M_{12}}f}{1 + P_{12}f} = \frac{t_1}{t_2}\frac{(1 + k)}{k} \tag{27}$$

Region I—Complete initial front series for $1 < Q < 1/k$. Within this region, if $M_{12}^2 > \dfrac{(1 - M_{12})k^2}{1 + P_{12}f}$, the maximum rebound occurs during the first bounce and

$$2CY_1 = \frac{(1 - M_{12})k^2}{1 + P_{12}f} \tag{28}$$

If the maximum rebound occurs later, it must occur during a complete heel series which follows the initial complete front series. From Equations (21) and (26)

$$2CY_1 = M_{12}^2 \tag{29}$$

By comparing Equations (28) and (29), the critical requirement

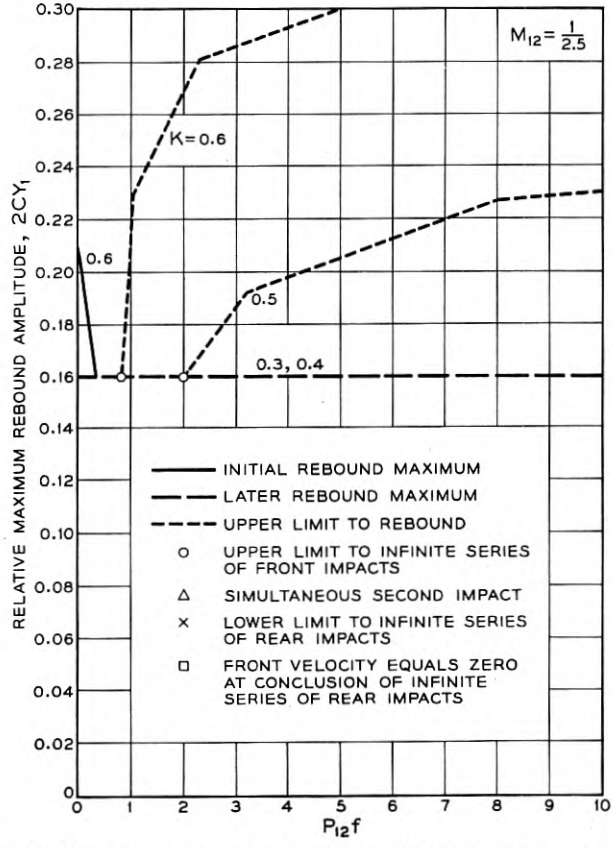


Fig. 5—Relative maximum rebound amplitude for $M_{12} = 1/2.5$.

for the latter case is that $P_{12}f > \frac{(1 - M_{12})}{M_{12}^2} k^2 - 1$. It should be noted that while the first rebound maximum, shown in solid lines on Figs. 5 to 10, is always realized, the later rebound given by (29) is an upper limit—shown in dashed lines—and is not always realized. In the dashed regions, phasing is extremely critical and small variations in the parameters may cause large variations in maximum rebound. From an engineering standpoint these regions are essentially undesirable.

Region II—Partial initial front series for

$$\frac{1}{k} < Q < \frac{1 + k}{k}.$$

This region is one of critical phasing, and attention is limited to special cases leading to maximum rebound. These cases occur when a

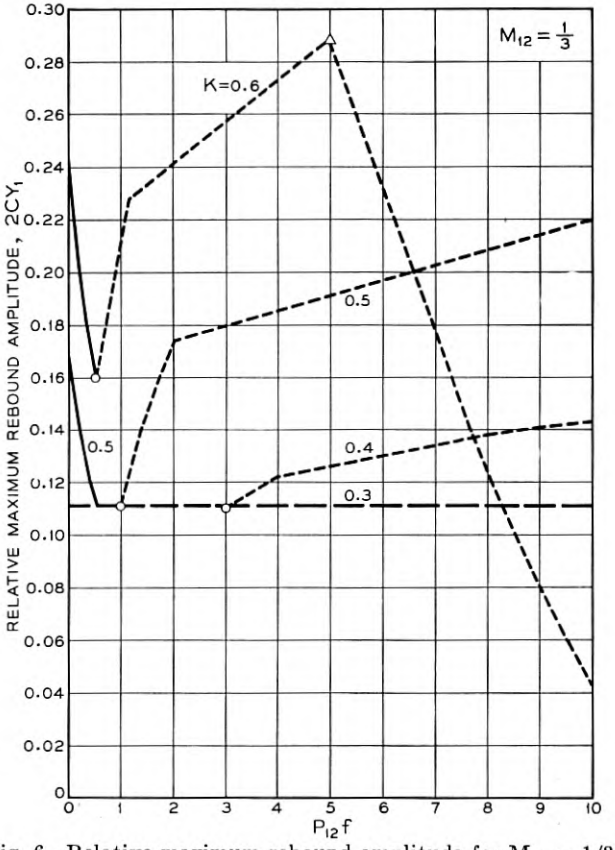


Fig. 6—Relative maximum rebound amplitude for $M_{12} = 1/3$.

heel impact immediately follows the last front impact of the series. These cases occur at

$$Q = \frac{1 - k^n}{k - k^n} \tag{30}$$

and lead to rebound amplitudes

$$2CY_1 = M_{12} + (1 - M_{12})[k^{2n} - M_{12}(1 - k^n)^2] \tag{31}$$

In Figs. 5 to 10, these special points are plotted and connected with straight dotted lines, which therefore indicate upper limits to rebound.

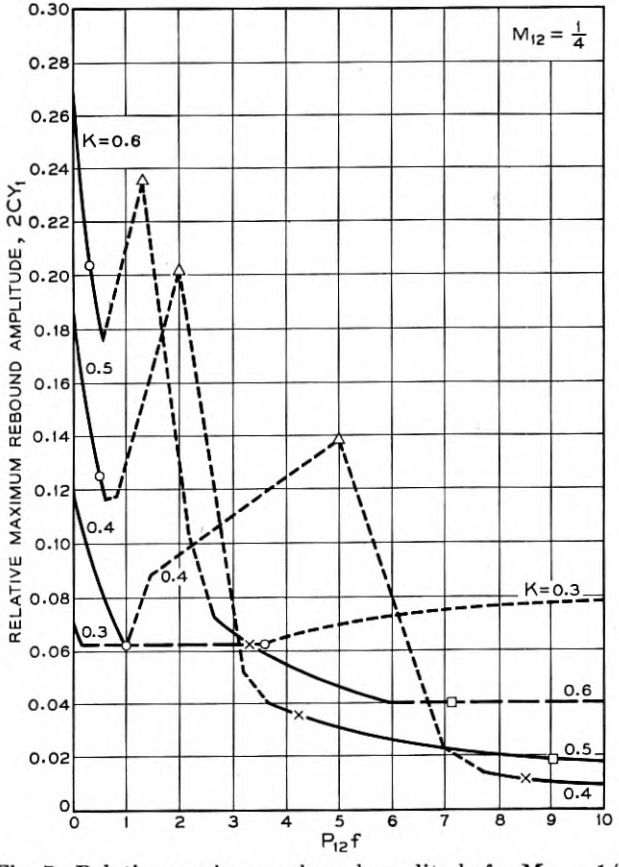


Fig. 7—Relative maximum rebound amplitude for $M_{12} = 1/4$.

Region III—Partial initial heel series for

$$\frac{1+k}{k} < Q < \frac{1+k}{k(1-k)+M_{12}k(1+k)}.$$

This is a region of critical phasing, and values were determined only for the maximum cases, where a front impact just precedes the last impact of the partial initial heel series. Here:

$$Q = \frac{1-k^{n+1}}{\frac{k(1-k)}{1+k} + k(1-k^n)M_{12}} \tag{32}$$

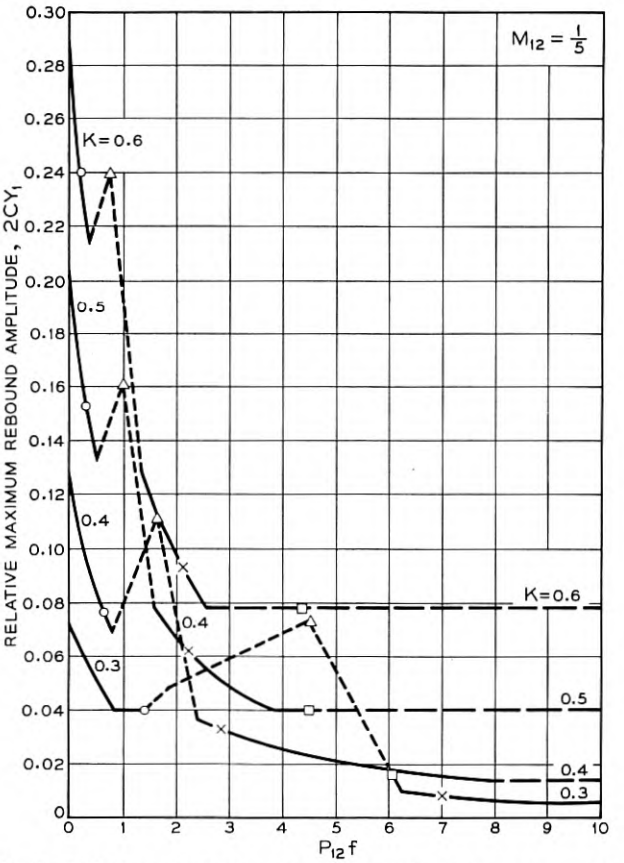


Fig. 8—Relative maximum rebound amplitude for $M_{12} = 1/5$.

and:

$$2CY_1 = 1 - (1 - M_{12})(1 - k^2)\{1 + [k - M_{12}(1 + k)(1 - k^n)]^2\} - M_{12}(1 - M_{12})(1 + k)^2\{1 - k^{2n} + [k - k^n - M_{12}(1 + k)(1 - k^n)]^2\} \quad (33)$$

Region IV—Complete initial heel series. A complete initial heel series implies that when the heel has come to rest, the front is still away from the stop. When the front finally meets its stop, the situation is identical with that just prior to the zero impact except that the energy content of the system is lower. The pattern must then repeat with diminished amplitudes. For this region we recognize two groups of cases. The first

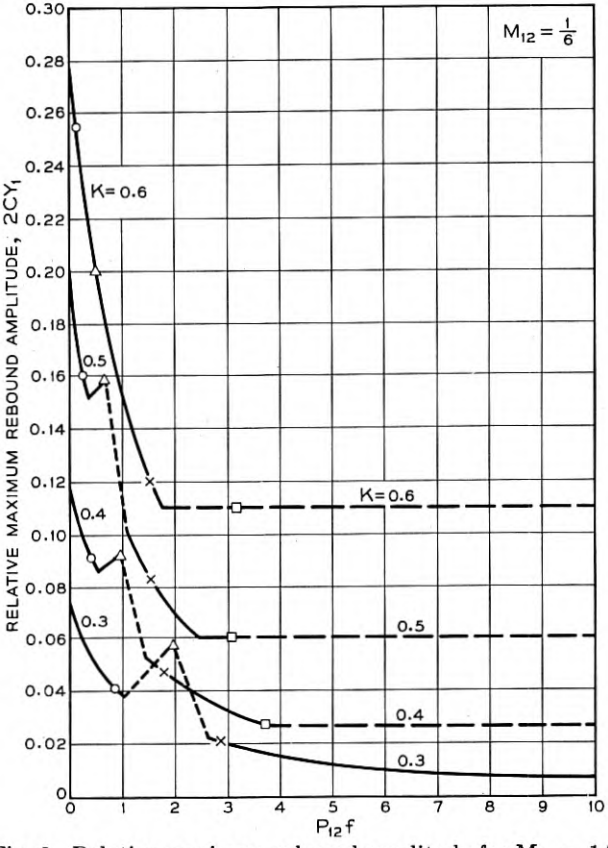


Fig. 9—Relative maximum rebound amplitude for $M_{12} = 1/6$.

group is that in which the front velocity is positive at the completion of the heel series. In that case

$$\frac{2(1+k)}{k(1-k)+M_{12}(1+k)^2} > Q > \frac{(1+k)}{k(1-k)+M_{12}k(1+k)^2}$$

and

$$2CY_1 = \frac{(1-M_{12})k^2}{1+P_{12}f} \tag{34}$$

For the second group the front velocity is still negative when the heel comes to rest from which point on the system acts as a one-degree-of-

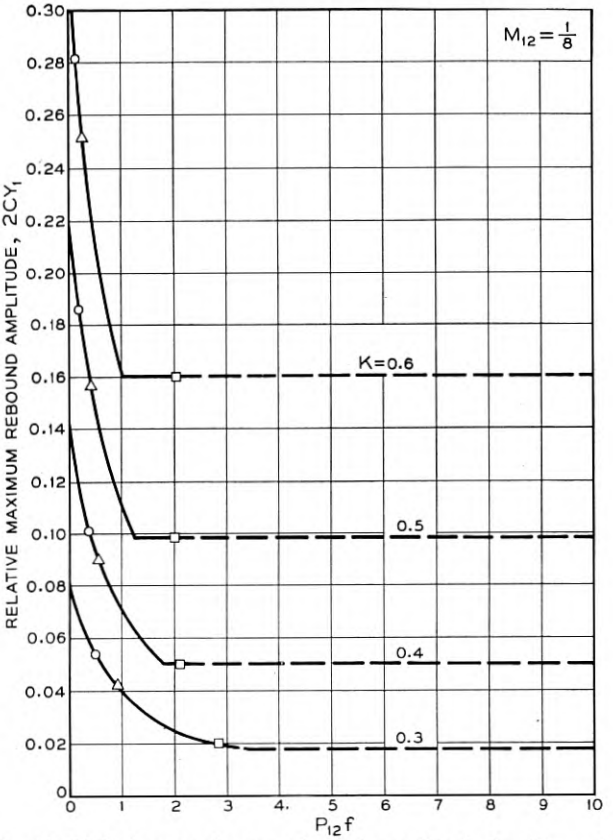


Fig. 10—Relative maximum rebound amplitude for $M_{12} = 1/8$.

freedom system until the next front impact. The requirement for this group is that

$$Q > \frac{2(1 + k)}{k(1 - k) + M_{12}(1 + k)^2}$$

and the maximum rebound is given by

$$2CY_1 = M_{12} - (1 - M_{12})[(1 - k^2) + M_{12}(1 + k)^2] \qquad (35)$$

It is to be noted that in the upper part of Group 1 the amplitude increases with successive heel impacts. This can be explored through the use of Equation (22). For simplicity of mapping, however, the limit given by Equation (35) has been extended back from the lower boundary of Group 2 until it intersects the line marking the first rebound amplitude of Group 1.

In Figs. 5 to 10 the respective regions have been identified by means of the symbols indicated below:

Region I	from $P_{12}f = 0$	to ○
Region II	from ○	to △
Region III	from △	to ×
Region IV, Group 1	from ×	to □
Region IV, Group 2	from □	to $P_{12}f \rightarrow \infty$

E. *Discussion of Rebound Charts*

Aside from quantitative data contained in Figs. 5 to 10, the following general trends are of interest:

For values of $M_{12} > \frac{1}{4}$, and the values of k under consideration, most of the useful range of $P_{12}f$ involves critical phasing and the rebound maxima are relatively high.

For values of $\frac{1}{6} \leq M_{12} \leq \frac{1}{4}$, consistently controllable rebound amplitude may be obtained.

For values of $M_{12} < \frac{1}{6}$ rebound increases again and the structure approaches the one-degree-of-freedom case.

VII. ANALYSIS OF REBOUND PATTERNS—THREE-DEGREES-OF-FREEDOM SYSTEM

Rebound pattern analysis as in Parts V and VI has so far not been performed for the three-degree-of-freedom system, partly because of complexity, and partly because for the system of Fig. 3 friction at the hinging stop will greatly influence the motion.

However, it is felt that the approach and notation of the analysis presented here is sufficiently general to allow extension of the rebound pattern analysis to the three-degree-of-freedom case. At any rate, with the assumption of the magnitudes of frictional forces, the basic equations of Part IV may be used to plot any particular case.

VIII. ARMATURE REBOUND MODEL

In order to verify the formal analysis presented in Parts III, IV and V, a large model of a two-degree-of-freedom system was constructed. It consisted essentially of a large bar constrained to move in a plane, biased against two stops, and to the ends of which writing pens were attached. As rebound conditions were simulated by releasing the bar against its stops, chart paper moved at right angles to the bar motion and thus produced a record of end displacement versus time.

By changing spring members and attaching masses to the bar, it was possible to vary the mass distribution and the biasing forces.

The results obtained closely agreed with those suggested by the analysis. The maximum rebound amplitudes were generally somewhat lower probably due to frictional effects.

IX. RELAY DESIGN CRITERIA RESULTING FROM ARMATURE REBOUND ANALYSIS

A. *Limitation of Analysis*

The assumptions which this analysis is subject to have been described under Part II. As applied to relays and probably the majority of mechanical structures, one assumption is most frequently and severely violated. The stops, which have been assumed to be stiff springs associated with a definite coefficient of restitution are, in practice, massive bodies which dissipate energy through excitation of high frequency modes of vibration. Accordingly, the assumption that the stops are at rest is violated, particularly if the mechanism is subject to repetitive (pulsing) impacts and the stop vibration does not decay greatly in the repetition period.

However, mechanisms designed in accordance with this analysis have performed well even under moderate pulsing conditions if the sensitive phasing region was avoided. In addition, every effort should be made to reduce the amount and duration of stop and mounting structure vibration by making them stiff, massive, and dissipative, if possible.

B. *Design Criteria*

1. Type of Armature Structure.

The selection of the number of degrees-of-freedom for an armature structure depends on the expected coefficient of rebound as well as practical considerations.

It can be shown without great difficulty that for very low coefficients of rebound the one-degree-of-freedom system is preferable. This is quite obvious when one considers the limiting value of $k = 0$. In this case the one-degree-of-freedom system will have no rebound whatsoever, while the two-degree-of-freedom system has a heel bounce followed by rebound at the front. The value of k below which the one degree system is preferable varies with the mass distribution relative to the stop points, being 0.18 for a rectangular plate armature with stops located at its edges.

Experience indicates that k in most practical relays and similar mechanical structures varies from 0.3 to 0.6. Hence the two-degree-of-freedom system is superior in all practical cases to the solidly hinged armature.

As far as three and higher degree-of-freedom systems are concerned, it may be said that generally the greater the number of modes resulting in impacts, the quicker the rebound energy can be diverted and dissipated and the lower theoretical rebound values can be obtained. This consideration would favor systems containing many degrees of freedom. However, while multi-degree-of-freedom systems can reach very low rebound values, their motion (phasing) must be very closely controlled or they may prove to be inferior to simpler systems particularly under vibratory (pulsing) operation. It is this difficulty which makes it appear that the two-degree-of-freedom system offers the best promise with the three-degree system also quite promising. By the same reasoning, additional spurious rocking modes should be minimized.

2. Armature Mass.

The armature mass should be as low as possible. This will minimize stop and structure vibration. In addition, in relay applications light armatures tend to increase magnetic "drag" losses of the armature during the release motion.

3. Stops and Mounting Structure.

As discussed before, it is desirable to reduce the amount and duration of stop and mounting structure vibration.

4. Coefficient of Restitution.

The coefficient of restitution should be kept low. Stops having low stiffness should, therefore, be avoided.

5. Biasing Forces.

F_1 should be kept as high as practicable.

For proper energy loss during impacts, the motion between impacts must occur outside the region of the compression, i.e., the armature and stop must separate. Therefore, because all practical stops have a finite stiffness, the biasing forces (F_1, F_2, etc.) should produce a static deflection less than say, arbitrarily, 5 per cent of the maximum expected rebound amplitude.

6. Design Parameters for Two-Degree-of-Freedom Systems.

As clearly indicated in Figs. 5 to 10 for the practical range of coefficients of restitution, most consistently good results are obtained with a coupling factor $M_{12} = \frac{1}{16}$ to $\frac{1}{4}$. This factor is most easily adjusted by correct placement of the front stop.

For the above range of M_{12} the force ratio F_2/F_1 should be such as to make the product

$$P_{12}\frac{F_2}{F_1} > 4 \qquad M_{12} = \tfrac{1}{4}$$

$$> 3 \qquad M_{12} = \tfrac{1}{5}$$

$$> 3 \qquad M_{12} = \tfrac{1}{6}$$

(Note: For a rectangular armature structure with the stops placed at its edges $M_{12} = \frac{1}{4}$, $P_{12} = \frac{1}{2}$ and force ratios in the neighborhood of 8 are desirable.)

X. ACKNOWLEDGMENT

The analytical treatment presented in this paper contains contributions by E. L. Norton, R. L. Peek, Jr., and the writer.

Appendix I

DERIVATION OF BASIC EQUATIONS OF MOTION THREE-DEGREE-OF-FREEDOM SYSTEM

(1) *Free Interval*

The motion of the armature will be described by the displacement at the stop points, x_1, x_2, x_3. Let m be the mass and R the radius of gyration of the armature about the center of gravity. The latter is located by the dimensions $\ell_1 R$, $\ell_2 R$, and $\ell_3 R$ relative to the stop points (Fig. 3).

The rotation and displacement of the center of gravity is then

$$\left.\begin{aligned} x_h &= (x_2 - x_1)\frac{\ell_3}{\ell_1 + \ell_2} + x_3 \\ x_v &= x_1 + (x_2 - x_1)\frac{\ell_1}{\ell_1 + \ell_2} \\ \theta &= \frac{x_2 - x_1}{R(\ell_1 + \ell_2)} \end{aligned}\right\} \quad (a)$$

From this the kinetic energy may be computed

$$\left.\begin{aligned} T &= \tfrac{1}{2}m(\dot{x}_h^2 + \dot{x}_v^2) + \tfrac{1}{2}mR^2\dot{\theta}^2 \\ &= \frac{\dot{x}_1^2(1 + \ell_2^2 + \ell_3^2) + \dot{x}_2^2(\ell_1^2 + \ell_3^2 + 1) + \dot{x}_3^2(\ell_1 + \ell_2)^2}{2(\ell_1 + l_2)^2} \\ &+ \frac{2\dot{x}_1\dot{x}_2(\ell_1\ell_2 - \ell_3^2 - 1) - 2\dot{x}_3\ell_3(\ell_1 + \ell_2)(\dot{x}_1 - \dot{x}_2)}{2(\ell_1 + \ell_2)^2} \end{aligned}\right\} \quad (b)$$

Applying LaGrange's Equation to the above, the equations of motion are obtained:

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_r}\right) - \frac{\partial T}{\partial q_r} = P_r$$

$$\left.\begin{aligned} \frac{F_1}{m} &= \frac{\ddot{x}_1[\ell_2^2 + \ell_3^2 + 1] + \ddot{x}_2[\ell_1\ell_2 - \ell_3^2 - 1] - \ddot{x}_3\ell_3(\ell_1 + \ell_2)}{(\ell_1 + \ell_2)^2} \\ \frac{F_2}{m} &= \frac{\ddot{x}_1[\ell_1\ell_2 - \ell_3^2 - 1] + \ddot{x}_2[\ell_1^2 + \ell_3^2 + 1] + \ddot{x}_3\ell_3(\ell_1 + \ell_2)}{(\ell_1 + \ell_2)^2} \\ \frac{F_3}{m} &= \frac{-\ddot{x}_1\ell_3(\ell_1 + \ell_2) + \ddot{x}_2\ell_3(\ell_1 + \ell_2) + \ddot{x}_3(\ell_1 + \ell_2)^2}{(\ell_1 + \ell_2)^2} \end{aligned}\right\} \quad (c)$$

The Equations (3) may be solved for $\ddot{x}_1$, $\ddot{x}_2$, $\ddot{x}_3$ and the results integrated, yielding

$$\left.\begin{aligned} x_1 &= \frac{1}{2m}[C_{11}F_1 + C_{12}F_2 + C_{13}F_3]t^2 + \dot{x}_{10}t + x_{10} \\ x_2 &= \frac{1}{2m}[C_{21}F_1 + C_{22}F_2 + C_{23}F_3]t^2 + \dot{x}_{20}t + x_{20} \\ x_3 &= \frac{1}{2m}[C_{31}F_1 + C_{32}F_2 + C_{33}F_3]t^2 + \dot{x}_{30}t + x_{30} \end{aligned}\right\} \quad (d)$$

where

$$\left.\begin{aligned} C_{11} &= (\ell_1^2 + 1) \qquad C_{13} = C_{31} = \ell_1\ell_3 \\ C_{22} &= (\ell_2^2 + 1) \qquad C_{12} = C_{21} = (1 - \ell_1\ell_2) \\ C_{33} &= (\ell_3^2 + 1) \qquad C_{23} = C_{32} = -\ell_2\ell_3 \end{aligned}\right\} \tag{3}$$

$\dot{x}_{10}$, $\dot{x}_{20}$, $\dot{x}_{30}$ are the initial velocities, x_{10} , x_{20} , x_{30} the initial displacements for the free interval in question. Interpretation of the analytic results is simplified by the introduction of normalization. Let $\dot{x}_a$ be $\dot{x}_1$ just before the "zero" impact and define

$$\left.\begin{aligned} y_i &= \frac{x_i}{\dot{x}_a\tau} = \frac{F_1}{\dot{x}_a^2 m}\, x_i \qquad \dot{y}_i = \frac{d}{d\left(\frac{t}{\tau}\right)}\, y_i = \frac{\dot{x}_i}{\dot{x}_a} \\ \tau &= \frac{\dot{x}_a m}{F_1} \end{aligned}\right\} \tag{2}$$

Dividing Equations (d) by $\dot{x}_a\tau$ yields the normalized equations of motion:

$$\left.\begin{aligned} y_1 &= \frac{1}{2}\left[C_{11} + C_{12}\frac{(F_2)}{(F_1)} + C_{13}\frac{(F_3)}{(F_1)}\right]\left(\frac{t}{\tau}\right)^2 + \dot{y}_{10}\left(\frac{t}{\tau}\right) + y_{10} \\ y_2 &= \frac{1}{2}\left[C_{21} + C_{22}\frac{(F_2)}{(F_1)} + C_{23}\frac{(F_3)}{(F_1)}\right]\left(\frac{t}{\tau}\right)^2 + \dot{y}_{20}\left(\frac{t}{\tau}\right) + y_{20} \\ y_3 &= \frac{1}{2}\left[C_{31} + C_{32}\frac{(F_2)}{(F_1)} + C_{33}\frac{(F_3)}{(F_1)}\right]\left(\frac{t}{\tau}\right)^2 + \dot{y}_{30}\left(\frac{t}{\tau}\right) + y_{30} \end{aligned}\right\} \tag{1}$$

(2) *Impact Interval*

The change of velocity at point "i" due to an impact at "i" is, by definition of the coefficient of restitution "k":

$$\Delta\dot{x}_i = -(1 + k_i)\dot{x}_i \tag{e}$$

Since this velocity change occurs as rotation about the conjugate point as an instant center of rotation, the impact relationships may be written, for an impact at point "1",

$$\Delta\dot{x}_1 = -(1 + k_1)\dot{x}_1$$

$$\Delta\dot{x}_2 = -(1 + k_1)\dot{x}_1 \frac{\left(\frac{R}{\ell_1} - \ell_2 R\right)}{\left(\ell_1 R + \frac{R}{\ell_1}\right)}$$

$$= (1 + k_1)\dot{x}_1 \frac{(\ell_1\ell_2 - 1)}{(\ell_1^2 + 1)} = -\frac{C_{12}}{C_{11}}(1 + k_1)\dot{x}_1$$

$$\Delta\dot{x}_3 = -(1 - k_1)\dot{x}_1 \frac{\ell_3 R}{\ell_1 R + \frac{R}{\ell_1}}$$

$$= -(1 + k_1)\dot{x}_1 \frac{\ell_1\ell_3}{\ell_1^2 + 1)} = -\frac{C_{13}}{C_{11}}(1 + k_1)\dot{x}_1$$

Similarly it can be shown that impacts at points (2) and (3) follow the same pattern. The general impact relations for impact at point "i" are then

$$\dot{y}_{j0n} = \dot{y}_{je(n-1)} + K_{ji}\dot{y}_{ie(n-1)} \tag{6}$$

The first subscript indicates the coordinate, and the second subscript indicates the beginning (0) or end (e) of the free interval denoted by the third subscript.

The impact transfer coefficient K_{ji} relating a velocity change at point "j" to an impact at point "i":

$$K_{ji} = -\frac{C_{ji}}{C_{ii}}(1 + k_i) \tag{7}$$

Appendix II

Analysis of Rebound Patterns—One-Degree-of-Freedom System

The equation of motion of this system is

$$y_{1n} = \tfrac{1}{2}Ct'^2 + \dot{y}_{10n}t' + y_{10n} \tag{f}$$

where

$$C = C_{11} - \frac{C_{12}^2}{C_{22}} \tag{9}$$

$$t' = \frac{t}{\tau}$$

and is measured from the start of the particular interval of free motion in question. The impact relationship is

$$\dot{y}_{10n} = -k_1\dot{y}_{1e(n-1)}$$

The motion consists of a series of parabolic arcs having periods of $2\dot{y}_{10}/C$ in general, or $2/C$, $2k/C$, $2k^2/C$, $\cdots$, $2k^{n-1}/C$. The time elapsed

is a convergent series and approaches, for a complete series:

$$\lim_{n\to\infty} \frac{2}{C}\left[1 + k + k^2 + \cdots k^n\right] = \frac{2}{C(1-k)} \tag{10}$$

The maximum rebound amplitude in any interval is $-\dot{y}_{10n}^2/2C$. The maximum excursion occurs during the first bounce at $t' = 1/C$ and equals $-k^2/2C$.

Appendix III

Analysis of Rebound Patterns—Two-Degree-of-Freedom System

The equations of motion of this system are

$$\left.\begin{aligned} y_1 &= \tfrac{1}{2}At'^2 + \dot{y}_{10n}t' + y_{10n} \\ y_2 &= \tfrac{1}{2}Bt'^2 + \dot{y}_{20n}t' + y_{20n} \end{aligned}\right\} \tag{g}$$

$$\left.\begin{aligned} \text{where } A &= C_{11} + C_{12}f \\ B &= C_{12} + C_{22}f \\ f &= \frac{F_2}{F_1} \\ t' &= \frac{t}{\tau} \text{ measured from the start of the particular free interval in question.} \end{aligned}\right\} \tag{h}$$

A. Complete Front Series

At the beginning of a front series

$$\left.\begin{aligned} y_1 &= 0 \\ \dot{y}_1 &= \dot{y}_{1c0} \\ y_2 &= y_{2c0} \\ \dot{y}_2 &= \dot{y}_{2c0} \end{aligned}\right\} \tag{i}$$

In a manner analogous to that for the one-degree-of-freedom system each front impact reduces $\dot{y}_1$ to $-k_1\dot{y}_1$. Therefore, after the n^{th} impact,

$$\dot{y}_{10n} = -k_1^n\dot{y}_{1c0}$$

and the time elapsed in the n^{th} interval is

$$T_n = \frac{2k_1^n}{A}\dot{y}_{1c0} \tag{j}$$

At the heel, from (g), the heel velocity preceding the n^{th} impact is

$$\dot{y}_{2e(n-1)} = \dot{y}_{20(n-1)} + Bt' \tag{k}$$

The velocity change during the $(n-1)$ interval is then equal to BT_{n-1}. From Equations (6), (7) and (12), the change in velocity during the n^{th} impact is $-P_{12}(1+k_1)k_1^{n-1}\dot{y}_{1e0}$.

The total change of $\dot{y}_2$ between impacts is then

$$\dot{y}_{20n} - \dot{y}_{20(n-1)} = BT_{n-1} - P_{12}(1+k_1)k_1^{n-1}\dot{y}_{1e0}$$

Similarly in preceding intervals:

$$\begin{aligned}
\dot{y}_{20(n-1)} - \dot{y}_{20(n-2)} &= BT_{n-2} - P_{12}(1+k_1)k_1^{n-2}\dot{y}_{1e0} \\
&\cdots \\
\dot{y}_{202} - \dot{y}_{201} &= BT_1 - P_{12}(1+k_1)k_1\dot{y}_{e0} \\
\dot{y}_{201} - \dot{y}_{2e0} &= - P_{12}(1+k_1)\dot{y}_{1e0}
\end{aligned}$$

By addition of the above

$$\begin{aligned}
\dot{y}_{20n} - \dot{y}_{2e0} &= B\sum_{m=1}^{n-1} T_m - P_{12}(1+k_1)\sum_{m=0}^{n-1} k_1^m\dot{y}_{1e0} \\
&= \frac{2B}{A}\dot{y}_{1e0}\sum_{m=1}^{n-1} k_1^m - P_{12}(1+k_1)\dot{y}_{1e0}\sum_{m=0}^{n-1} k^m
\end{aligned}$$

The summations may be evaluated, yielding

$$\dot{y}_{20n} - \dot{y}_{2e0} = \left[\frac{2B}{A}\frac{k_1 - k_1^n}{1-k_1} - P_{12}(1+k_1)\frac{1-k_1^n}{1-k_1}\right]\dot{y}_{1e0} \tag{l}$$

To evaluate the displacements at the heel, Equation (g) yields

$$y_{20n} - y_{20(n-1)} = \dot{y}_{20(n-1)}T_{n-1} + \tfrac{1}{2}BT_{n-1}^2$$

Adding these expressions for intervals 0 to n; the total change in y_2 is

$$\begin{aligned}
y_{20n} - y_{201} &= \sum_{m=1}^{n-1}\dot{y}_{20m}T_m + \tfrac{1}{2}B\sum_{m=1}^{n-1}T_m^2 \\
&= \frac{2k_1(1-k_1^{n-1})}{A(1-k_1)}\dot{y}_{1e0}\dot{y}_{2e0} \\
&+ \left[\frac{2B(k_1^2 - 2k_1^{n+1} + k_1^{2n})}{A^2(1-k_1)^2}\right. \\
&\left. - \frac{2P_{12}k_1(1-k_1^n - k_1^{n-1} + k_1^{2n-1})}{A(1-k_1)^2}\right]\dot{y}_{1e0}^2
\end{aligned} \tag{m}$$

Expressions for an initial series may be obtained by setting $\dot{y}_{1e0} = 1$ $\dot{y}_{2e0} = y_{2e0} = 0$, and, finally, for an initial complete series $m \to \infty$ and hence $k^m \to 0$, and Equations (l) and (m) become

$$\left.\begin{aligned} y_{2e\infty} &= \frac{2k_1}{A(1+k_1)^2}\left[\frac{Bk_1}{A} - P_{12}\right] \\ \dot{y}_{2e\infty} &= \frac{1}{1-k_1}\left[\frac{2Bk_1}{A} - P_{12}(1+k_1)\right] \end{aligned}\right\} \tag{17}$$

B. Complete Heel Series

For heel series, Equations (l) and (m) may be used by interchanging the initial velocities, accelerations, and impact transfer coefficients for those relating to heel motion:

$$\dot{y}_{10n} - \dot{y}_{1e0} = \left[\frac{2A}{B}\frac{k_2 - k_2^n}{1-k_2} - \frac{M_{12}(1+k_2)}{P_{12}}\frac{1-k_2^n}{1-k_2}\right]\dot{y}_{2e0} \tag{n}$$

$$\begin{aligned} y_{10n} - y_{101} &= \frac{2k_2(1-k_2^{n-1})}{B(1-k_2)}\dot{y}_{1e0}\dot{y}_{2e0} \\ &+ \left[\frac{2A(k_2^2 - 2k_2^{n+1} + k_2^{2n})}{B^2(1-k_2^2)} - \frac{2M_{12}k_2(1-k_2^n - k_2^{n-1} + k_2^{2n-1})}{BP_{12}(1-k_2^2)}\right]\dot{y}_{2e0}^2 \end{aligned} \tag{o}$$

An initial heel series occurs when the heel strikes first after the "zero" impact. The first heel impact then occurs $T_1 = 2P_{12}/B(1+k_1)$ after the zero impact and the initial conditions are

$$\dot{y}_{2e1} = P_{12}(1+k_1)$$

$$\dot{y}_{1e1} = -k_1 + AT_1 = \frac{2AP_{12}}{B}(1+k_1) - k_1$$

$$y_{1e1} = -k_1T_1 + \tfrac{1}{2}AT_1^2 = \frac{2P_{12}}{B}(1+k_1)\left[\frac{AP_{12}}{B}(1+k_1) - k_1\right]$$

Substitution of the above into (n) yields

$$\dot{y}_{10n} = -k_1 + \frac{1+k_1}{1-k_2}\left[\frac{2AP_{12}}{B}(1-k_2)^n - M_{12}(1+k_2)(1-k_2^n)\right] \tag{p}$$

The corresponding expression for y_{10n} is quite involved. For the special case of $k = k_1 = k_2$

$$y_{10n} = \frac{AP_{12}(1+k)^2}{B}\left[\frac{AP_{12}}{B}\left(\frac{1-k^n}{1-k}\right)^2 = \frac{k(1-k^n)}{1-k^2} - \frac{M_{12}(1-k^n)(k-k^n)}{(1-k)^2}\right] \tag{q}$$

If the initial series is a complete series, $n \to \infty$ and

$$\left.\begin{aligned} y_{10\infty} &= \frac{AP_{12}(1+k)^2}{B(1-k)^2}\left[\frac{AP_{12}}{B} - \frac{k(1-k)}{1+k} - M_{12}k\right] \\ \dot{y}_{1e\infty} &= \frac{1+k}{1-k}\left[\frac{2AP_{12}}{B} - \frac{k(1-k)}{(1+k)} - M_{12}(1+k)\right] \end{aligned}\right\} \tag{22}$$

C. Partial Front Series

The worst rebound occurs when heel and front impacts occur nearly simultaneously, with the front hitting first. From Equation (m) for an initial front series, this requires that

$$\frac{B}{AP_{12}} = Q = \frac{1-k^n}{k-k^n} \tag{30}$$

After n front impacts conditions are given by Equations (14) and (19) with $y_1 = y_2 = 0$, and

$$\frac{T+V}{T_0} = 1 - (1-M_{12})(1-k^{2n}) = k^{2n} + \frac{M_{12}\dot{y}_2^2}{P_{12}^2} - \frac{2M_{12}k^n\dot{y}_2}{P_{12}}$$

This may be solved for $\dot{y}_2 = P_{12}(1-k^n)$. The maximum front excursion now possible is that for a complete series of heel impacts. The above value of $\dot{y}_2$ in Equation (24) yields

$$2CY_1 = M_{12} + (1-M_{12})[k^{2n} - M_{12}(1-k^n)^2] \tag{31}$$

D. Partial Heel Series

The worst rebound occurs again when heel and front impacts occur nearly simultaneously, with the front hitting first. From Equation (9) for an initial heel series, this requires that

$$\frac{B}{AP_{12}} = Q = \frac{1-k^{n+1}}{\dfrac{k(1-k)}{1+k} + k(1-k^n)M_{12}} \tag{32}$$

After n heel impacts $\dot{y}_2 = P_{12}(1 + k)k^n$ and from Equations (19) and (23)

$$\frac{T + V}{T_0} = 1 - (1 - M_{12})(1 - k^2) - M_{12}(1 - M_{12})(1 + k)^2(1 - k^{2n})$$
$$= \dot{y}_1^2 - 2M_{12}(1 + k)k^n\dot{y}_1 + M_{12}(1 + k)^2k^{2n}$$

This may be solved for $\dot{y}_1 = k - M_{12}(1 + k)(1 - k^n)$ and after the front impact immediately following:

$$\dot{y}_2 = P_{12}(1 + k)k^n - P_{12}(1 + k)[k - M_{12}(1 + k)(1 - k^n)]$$

The maximum front excursion now possible is that for a complete series of heel impacts. The above value for $\dot{y}_2$ in Equation (24) yields

$$2CY_1 = 1 - (1 - M_{12})(1 - k^2)\{1 + [k - M_{12}(1 + k)(1 - k^n)]^2\}$$
$$- M_{12}(1 - M_{12})(1 + k)^2\{1 - k^{2n} \qquad (33)$$
$$+ [k - k^n - M_{12}(1 + k)(1 - k^n)]^2\}$$

Appendix IV

SUMMARY OF NOTATION

$A = C_{11} + C_{12} + C_{12}f$

$B = C_{12} = C_{22} + C_{22}f$

$C = C_{11} - \dfrac{C_{12}^2}{C_{22}}$

$C_{11} = 1 + \ell_1^2 \qquad C_{12} = C_{21} = [1 - \ell_1\ell_2]$

$C_{22} = 1 + \ell_2^2 \qquad C_{13} = C_{31} = \ell_1\ell_3$

$C_{33} = 1 + \ell_3^2 \qquad C_{23} = C_{32} - \ell_2\ell_2$

$f = \dfrac{F_2}{F_1}$

F_1 = front tensioning force

F_2 = heel tensioning force

k_1 = coefficient of restitution at vertical front stop

k_2 = coefficient of restitution at vertical heel stop

$$K_{ji} = -\frac{C_{ji}}{C_{ii}}(1 + k_i)$$

$\ell_1 R$ = vertical front stop location relative to c.g.

$\ell_2 R$ = vertical heel stop location relative to c.g.

$\ell_3 R$ = horizontal heel stop location relative to c.g.

m = mass of armature

$$M_{ij} = \frac{C_{ij}^2}{C_{ii}C_{ij}}$$

$$P_{ij} = \frac{C_{ij}}{C_{ii}}$$

$$Q = \frac{1 + \dfrac{P_{12}}{M_{12}f}}{1 + P_{12}f} = \frac{t_1}{t_2}\frac{1 + k_1}{k_1}$$

R = radius of gyration of armature about center of gravity

$$\tau = \frac{\dot{x}_a m}{F_1}$$

t = time

$$t' = \frac{t}{\tau}$$

t_1 = basic period of front after "zero" impact

t_2 = basic period of heel after "zero" impact

T_n = duration of n^{th} free interval

x_1 = vertical front displacement

x_2 = vertical heel displacement

x_3 = horizontal heel displacement

$\dot{x}_a$ = front velocity just prior to "zero" impact

$$y = \frac{x}{\dot{x}_a \tau}$$

Y_1 = greatest excursion (rebound) of front

Abstracts of Bell System Technical Papers Not Published in This Journal

A New Telephone Carrier System for Medium-Haul Circuits. R. S. CARUTHERS [1], H. R. HUNTLEY[2], W. E. KAHL[1], and L. PEDERSON[1]. *Elec. Eng.*, **70**, pp. 692–693, Aug. 1951.

Low terminal costs and single-cable operation make this the first economically practical carrier system for medium-haul telephone circuits. Performance is not sacrificed for economy.

*A .15-Kw 500-Megacycle Grounded-Grid Tridode.** C. E. FAY[1], D. A. A. HALE[1], and R. J. KIRCHER[1]. *Proc. I.R.E.*, **39**, pp. 800–803, July, 1951.

Short-Cut Method Aids Figuring Exhaust and Collecting Systems. W. H. FOGLE[3]. *Heating, Piping and Air Cond.*, **23**, pp. 75–78, July, 1951.

A simple, workable method of determining static losses in industrial exhaust and collecting systems is explained here by means of a sample problem and its step-by-step solution. The method might be described as an "averaging out" process, whereby all duct sizes and lengths are averaged out with the cfm and velocity to give a total linear footage of the average duct size and an average velocity.

*Arcing at Electrical Contacts on Closure. Part I. Dependence Upon Surface Conditions and Circuit Parameters.** L. H. GERMER[1]. *Jl. Appl. Phys.* **22**, pp. 955–964, July, 1951.

In a low-voltage circuit the occurrence of an arc between approaching electrodes is dependent upon the nature of the surfaces and upon the circuit inductance. For carbon surfaces, or noble metal surfaces which have been "activated" by operation in various organic vapors resulting in a carbonaceous layer, the limiting circuit inductance is somewhat above 10^{-3}h, which is much higher than the limiting inductance for clean noble metal surfaces. This activation by organic vapors occurs for noble metals only and for certain vapors; for example, benzene derivatives. In the case of silver and benzene vapor, it has been shown that the activation is due to adsorption of benzene onto a greasy surface layer and its decomposition there by the heat of subsequent closures. A metal surface, which has been activated by organic vapor, remains active indefinitely if there

* A reprint of this article may be obtained on request.
[1] Bell Tel. Labs.
[2] A. T. & T. Co.
[3] W. E. Co.

is no arcing at the surfaces; but with continued operation and accompanying arcing, the activating material is burned away, and the surface returns to the inactive condition if no activating vapor is supplied.

Arc voltages, which are independent of current and of ambient gas, as far as tested, have been measured for a number of metals and for carbon; the arc voltage for carbon is quite erratic in the range between 20 and 30 volts, but for each of a number of metals the arc voltage is steady.

Arcing at noble metal surfaces, similar to that induced by carbonaceous material from organic vapors, can be produced also by insulating particles or insulating films. The active condition gradually disappears with continued arcing, unless there is a steady supply of insulating material to the surface.

The minimum arc current has been measured to be 0.6 amp for active silver and for carbon, and 0.03 amp for inactive silver. These are the currents at which an established arc is extinguished.

*Iron-Silicon Alloys Heat Treated in a Magnetic Field.** M. GOERTZ[1]. *Jl. Appl. Phys.*, **22**, pp. 964–965, July, 1951.

Heat treatment in a magnetic field has been found effective for iron-silicon alloys between two per cent and ten per cent silicon, the highest maximum permeability being obtained at about 6.5 per cent silicon. In a single crystal of this composition, magnetized parallel to a (100) direction, the hysteresis loop is squared by the magnetic anneal and the maximum permeability is increased from 50,000 to 3,800,000, the highest value yet reported.

Domain Boundary Motion in Ferroelectric Crystals and the Di-electric Constant at High Frequency. C. KITTEL[1]. Letter to the Editor. *Phys. Rev.*, **83**, p. 458, July 15, 1951.

*A Method for Determining the Propagation Constants of Plastics at Ultrasonic Frequencies.** H. J. MCSKIMIN[1]. *J. Acoust. Soc. Am.*, **23**, pp. 429–434, July, 1951.

A pulse technique particularly suited to dissipative materials is described for measuring attenuation and phase-shift constants of plastics, using either transverse or longitudinal waves in the frequency range of 5–50 mc.

A thin wafer of the material under test is placed between two identical fused silica buffers; and waves generated by quartz crystals at the ends of the assembly are transmitted simultaneously through the specimen in both directions. Comparison of transmitted and reflected components by means of a special balancing circuit provides information from which the complex propagation constant can be calculated, and hence dynamic rigidities and viscosities.

Illustrative data for polyethylene and Nylon are given.

* A reprint of this article may be obtained on request.
[1] Bell Tel. Labs.

Some Antecedents of Quality Control. E. C. MOLINA[4]. *Ind. Quality Control*, **8**, pp. 10–11, July, 1951.

Traveling-Wave Amplifier Measurements. F. E. RADCLIFFE[1]. *Electronics*, **24**, pp. 110–111, Aug., 1951.

Rapid sweep-frequency technique used at 4,000 mc can be applied to all broad-band amplifier measurements. Oscilloscope display shows transmission accurate to 0.1 db and return-loss values up to 40 db.

*Kirchhoff's Formula, its Vector Analogue, and Other Field Equivalence Theorems.** S. A. SCHELKUNOFF[1]. *Communications on Pure and Applied Math.*, **4**, pp. 43–59, June, 1951.

*Remarks Concerning Wave Propagation in Stratified Media.** S. A. SCHELKUNOFF[1]. *Communications on Pure and Applied Math.*, **4**, pp. 117–128, June, 1951.

*An Achromatic Doublet of Silicon and Germanium.** R. G. TREUTING[1]. *J. Opt. Soc. Am.*, **41**, pp. 454–456, July, 1951.

The semi-metals germanium and silicon have high transparency and high refractive indices over a wide range of infrared wavelengths and are stable to normal atmospheres. Their relative indices and dispersions make achromatic combinations possible; and designs are given for axially corrected doublets of relative apertures f:2 and f:1. The optical homogeneity of the materials is discussed: compositional variations are not considered an optical hazard, but there is evidence of structural imperfections in some specimens whose effect on optical properties remains to be evaluated.

*On the Motion of Gaseous Ions in a Strong Electric Field. I.** G. H. WANNIER[1]. *Phys. Rev.*, **83**, pp. 281–289, July 15, 1951.

This paper applies the Boltzmann method of gaseous kinetics to the problem of positive ions moving through a gas under the influence of a static, uniform electric field. The ion density is assumed to be vanishingly low, but the field is taken to be strong; that is, the energy which it imparts to the ions is not assumed negligible in comparison to thermal energy. Attention is focused upon the computation of velocity averages, and the drift velocity in particular, rather than a complete knowledge of the entire velocity distribution. It is shown in Sections C and E that the problem so formulated is completely soluble if the mean free time between collisions of ions and molecules is a constant; this is the case for the so-called polarization force between ions and molecules which predominates over other forces at low temperature. A method for obtaining averages to any desired accuracy in the general case is developed in Section D. The method is applied to the hard sphere model for the high field range and

* A reprint of this article may be obtained on request.
[1] Bell Tel. Labs.
[4] Bell Tel. Labs., Retired.

mass ratio 1. An application of the resulting formula (43) to experimental material has been published earlier.

Evidence for the Noncubic High Temperature Phase of $BaTiO_3$. E. A. WOOD[1]. Letter to the Editor. *J. Chem. Phys.*, **19**, p. 976, July, 1951.

*Seven-League Oscillator.** F. B. ANDERSON[1]. *Proc. I.R.E.*, **39**, pp. 881–890, Aug., 1951.

A bridge-type RC oscillator is described which is continuously adjustable over a frequency range of 20 cps to 3 mc in one sweep of a two-gang linear potentiometer control. Tracking requirements of the two-gang control are not severe. The output is available in four phases, and the frequency is an approximately logarithmic function of the linear potentiometer setting. Practical limits of the frequency range are tentatively 0.01 cps and 10 mc. Accuracy of setting of the order of one per cent is attainable with ordinary components. Frequency stability is of the order of 2 per cent per db of tube gain variation.

*Semi-Conductor Surface Phenomena.** W. H. BRATTAIN[1]. *Semi-Conducting Materials*, H. K. HENISCH, ed., pp. 37–46. Proceedings of a conference held at the University of Reading (July 10–15, 1950) London, Butterworths, 1951.

Developments in the understanding and interpretation of phenomena occurring at the surface of a semi-conductor are reviewed. The development starts with the Mott-Schottky theory of the space charge layer. Bardeen's concept of a space charge layer due to 'surface states' explained the independence of rectification on contact potential and Meyerhoff's small values for the contact potential between n- and p-type silicon. Shockley and Brattain observed that this contact potential difference increased with impurity concentration. Illumination of a silicon surface produced hole and electron pairs in the space charge layer. The potential of the surface changed until the photocurrent was balanced by a conduction current. The relation between photo-current and potential change was of the same form as a forward characteristic for a rectifying contact. In an experiment similar to Becquerel's, using water as an electrolyte, the surface may be biased in the reverse direction. When so biased the response to modulated light gives the differential resistance of the space charge layer. This resistance increases rapidly with reverse bias and the time constant of the layer increases, both agreeing qualitatively with theory. This experimental method of measuring changes in surface potential caused by illumination permits determination of the properties of the space charge layer at the free surface of a semiconductor.

*The Calculation of Traveling-Wave-Tube Gain.** C. C. CUTLER[1]. *Proc. I.R.E.*, **39**, pp. 914–917, Aug., 1951.

* A reprint of this article may be obtained on request.
[1] Bell Tel. Labs.

Essential information for claculation of traveling-wave-tube gain is summarized and condensed in this paper. The important relations are documented, presented in a concise form for simplified computation, and developed as a nomograph. The conclusions have been found to be in agreement with measurements on six different tube designs.

*Thermal Variation of Young's Modulus in some Fe-Ni-Mo Alloys.** M. E. FINE[1] and W. C. ELLIS[1]. References. *Jl. Metals*, **3**, pp. 761–764, Sept., 1951.

*A Broad-Band Transcontinental Radio Relay System.** T. J. GRIESER[1] and A. C. PETERSON[1]. *Elec. Eng.*, **70**, pp. 810–815, Sept., 1951.

Spanning the continent from coast to coast, this microwave relay system provides six channels, each of which can carry one television circuit or hundreds of telephone circuits. Some features of this vast network are descirbed.

*An Improved Telephone Set.** A. H. INGLIS[1] and W. L. TUFFNELL[1]. *Elec. Eng.*, **70**, pp. 770–775, Sept., 1951.

The familiar telephone set has undergone numerous changes which will provide better service at lower cost than do present models. Increased transmitting and receiving gain, better sidetone control, broader frequency response, faster dialing, simple ringing control, and a trim appearance are some of the features of the new design.

The Institutes for Basic Research—Their Contribution to National Strength. M. J. KELLY[1], pp. 11–23. *Applied Research is Not Enough*, (booklet). Addresses at the Dedication of the Institutes for Basic Research, The University of Chicago, May 16, 1951.

The Crystal Clock. W. A. MARRISON[1]. *Science Marches On*, JAMES STOKLEY, ed., N. Y., Ives Washburn, Inc., 1951, pp. 303–309.

Observations of Zener Current in Germanium p-n Junctions. K. B. MCAFEE[1], E. J. RYDER[1], W. SHOCKLEY[1], and M. SPARKS[1]. Letter to the Editor. *Phys. Rev.*, **83**, pp. 650–651, Aug. 1, 1951.

*Experimental Radio-Telephone Service for Train Passengers.** N. MONK[1]. *Proc. I.R.E.*, **39**, pp. 873–881, Aug., 1951.

Experimental public radio-telephone service for train passengers was inaugurated by the Bell Telephone System several years ago. Initial installations operated in conjunction with a series of urban mobile base stations. More recently, highway mobile systems have been used for this service, and this paper describes a typical train installation operating through a highway channel. All

* A reprint of this article may be obtained on request.
[1] Bell Tel. Labs.

of these early systems utilized an attendant on the train. The cost of providing an attendant has, in some cases, been found excessive. Consequently, experiments have been initiated in which a coin box is used on the train. The arrangements for this purpose are also described.

*The Magneto-Resistance Effect in Oriented Single Crystals of Germanium.** G. L. PEARSON[1] and H. SUHL[1]. *Phys. Rev.*, **83**, pp. 768–776, Aug. 15, 1951.

This paper describes an extensive study of the magneto-resistance effect in germanium as a function of crystal orientation. Experimental measurements establish the constants involved in the dependence of the effect on orientation of magnetic field and electric relative to the crystal axes. The measurements are internally consistent with existing phenomenological theory based on cubic crystal symmetry, in which terms involving the magnetic field to higher than the second order are neglected. It is shown that such deviations as do occur arise from higher terms in the field, since an extension of the phenomenological theory to the fourth order predicts their symmetry. Relations are established between the experimentally observed phenomenological constants and those constants appearing in existing magneto-resistance electronic theories. It is concluded that no electronic theory yet worked out is entirely consistent with experiment. The present electronic theories are special cases of a very general theory recently proposed by Shockley, and it is possible that agreement can be obtained as soon as the computational difficulties of the latter theory are overcome.

New Phenomena of Electronic Conduction in Semi-Conductors. W. SHOCKLEY[1]. *Semi-Conducting Materials*, H. K. HENISCH, ed., pp. 26–36. Proceedings of a conference held at the University of Reading (July 10–15, 1950) London, Butterworths, 1951.

The semi-conductors silicon and germanium may be discussed as insulators the electronic structure of which is disturbed. Excess electrons, which act as negatively charged current carriers, may be present as may be 'holes' or places where electrons are missing from the valence-bond structure. Holes act as positively charged current carriers. In ordinary electronic conduction the flow of current carriers is substantially incompressible so that the density of carriers remains constant. When a new transistor phenomenon known as 'carrier injection' occurs, however, the total density of holes and electrons may be greatly increased and this modulation of the electronic structure may be used both for scientific investigation and for practical amplification. In particular, carriers may be injected at a predetermined time and place into a known uniform electric field and their transit time to another place accurately timed by detecting their arrival with a "collector" point. Drift velocities and mobilities may be measured precisely in this way with a directness unattainable by pre-transistor

* A reprint of this article may be obtained on request.
[1] Bell Tel. Labs.

methods involving conductivity and Hall effect. These new experiments and their related theories may furnish a foundation for a new engineering science of transistor electronics.

*Domain Patterns on Nickel.** H. J. WILLIAMS[1] and J. G. WALKER[1]. *Phys. Rev.*, **83**, pp. 634–636, Aug. 1, 1951.

Domain patterns have been observed on two single crystals of nickel cut in the form of hollow parallelograms. The length of the sides were parallel to the (111) directions in one specimen and to the (110) directions in the other. The crystals show domain structures with the three types of domain boundaries which are to be expected from a material having the directions of easy magnetization along the (111) directions. Domain boundary movement under the influence of an applied magnetic field was observed.

*Polymorphism in Potassium Niobate, Sodium Niobate, and Other ABO_3 Compounds.** E. A. WOOD[1]. Bibliography. *Acta Cryst.*, **4**, pp. 353–362, July, 1951.

The first part of this paper presents the results of optical and X-ray studies of the perovskite-type crystals, potassium niobate and sodium niobate. Potassium niobate is orthorhombic at room temperature, changing to tetragonal at about 225°C. and cubic near 435°C. Sodium niobate is orthorhombic at room temperature, changing to tetragonal at about 370°C. and to cubic at about 640°C.

The second part of the paper discusses relations among the structures of the ABO_3 compounds.

*Subjective Sharpness of Additive Color Pictures.** M. W. BALDWIN[1]. *Proc. I.R.E.*, **39**, pp. 1173–1176, Oct., 1951.

This is a report on the first numerical results to come from a laboratory experiment on the subjective sharpness of additive three-color pictures. The sharpness factor is isolated by using out-of-focus projection (of slides) instead of actual television transmission.

An observer's acuity for defocus is greatest for the green component and least for the blue component, in an additive three-color picture. When the same picture is reproduced in monochrome (white, red, green, or blue) at the same brightness, the observer's acuity for defocus is equal to that found for the green component.

*Frequency-Modulation Terminal Equipment for the Transcontinental Relay System.** J. G. CHAFFEE[1] and J. B. MAGGIO[1]. *Elec. Eng.*, **70**, pp. 880–883, Oct., 1951.

To meet the exacting requirements of the new transcontinental microwave relay system, specially designed frequency-modulation terminal equipment was

* A reprint of this article may be obtained on request.
[1] Bell Tel. Labs.

constructed. The terminal transmitter converts either message or television signals to a frequency-modulated signal centered on 70 mc and the terminal frequency-modulation receiver recovers these signals, thus providing a link between the relay system and other telephone facilities.

*Observer Reaction to Low-Frequency Interference in Television Pictures.** A. D. FOWLER[1]. *Proc. I.R.E.*, **39**, pp. 1332–1336, Oct., 1951.

This paper presents results of tests to determine how much low-frequency interference can be tolerated in black-and-white television pictures. Various levels of single low-frequency interference were superimposed on a locally transmitted television picture. Observers viewed the picture and rate the disturbing effect of each level of the interference. Ratings were made in terms of preworded comments ranging from "not perceptible" to "unusable." Interfering frequencies from 48 to 90 cycles per second were employed.

Just visible interference appears as a flicker. The rate of flicker is the difference between interfering and 60-cycle field frequencies. The most disturbing interference produced a flicker rate of 5 or 6 cycles per second. To be tolerated, peak-to-peak amplitude of this interference had to be 54 db weaker than the peak-to-peak amplitude of the television signal (including synchronizing pulse). For flicker rates of 0.5 and 12 cycles per second, the amount of interference which could be tolerated was larger by 14 and 3 db, respectively.

*Arcing at Electrical Contacts on Closure. Part II. The Initiation of an Arc.** L. H. GERMER[1]. *Jl. Appl. Phys.*, **22**, pp. 1133–1139, Sept., 1951.

The capacity of the plates of an oscilloscope charged to 35 or 40 volts is discharged repeatedly by approaching electrodes of carbon, active silver, and inactive silver. Facts about the discharges, which are arcs of very short duration, are inferred from resulting open circuit potentials and calculated electrode separations.

The separation at the first arc varies in different experiments but corresponds on the average to a nominal electric field of 0.6×10^6 volts/cm for carbon or active silver and to 2×10^6 volts/cm for inactive silver. Each arc is initiated by a very small number of field emission electrons. The hypothesis that a single electron may perhaps be sufficient is consistent with observations at later stages of each closure when the electrodes are closer and the field much higher.

The earlier observation, that the potential across a short arc is constant and independent of current, is not true if the arc time is sufficiently short. For active silver a time comparable with 2×10^{-8} sec is required to establish the steady arc voltage characteristic of later stages of arcs which last longer than this. The initial time during which the potential is decreasing toward its final steady value is 100 times the transit time of a silver ion across the gap.

---

* A reprint of this article may be obtained on request.
[1] Bell Tel. Labs.

Computation of Control Limits for p-Charts When the Samples Vary in Size. H. L. JONES[5]. *Ind. Quality Control*, **8**, pp. 26–27, Sept., 1951.

The Design of Switching Circuits. W. KEISTER[1], A. E. RITCHIE[1], and S. H. WASHBURN[1]. *N. Y., Van Nostrand*, 1951. 556 pp. (Bell Telephone Laboratories Series).

This is the first published textbook in its field. It presents, first, the fundamental design principles of switching circuits composed of discrete-valued switching elements. Most of the discussion concerns two-valued elements, with greatest emphasis placed on electromagnetic relays. Chapters cover basic circuit paths, the logical interpretation of requirements, and the techniques of constructing networks to fulfill these requirements. The symbolic methods of Boolean algebra and its application to the design of combinational and sequential circuits is covered. Later chapters cover various unifunctional circuits such as selecting, connecting, translating, counting, and lockout. Final chapters discuss methods of synthesising unifunctional circuit building blocks into larger circuits and systems.

Measurement of the Elastic Constants of Silicon Single Crystals and Their Thermal Coefficients. H. J. MCSKIMIN[1], W. L. BOND[1], E. BUEHLER[1], and G. K. TEAL[1]. Letter to the Editor. *Phys. Rev.*, **83**, p. 1080, Sept. 1, 1951.

Interest in the properties of silicon single crystals arising from their use as semiconductors has led us to make measurements of the elastic constants of two single crystals. Measurements of velocities of propagation for both shear and longitudinal waves were made in the crystals as described in a recent paper by McSkimin. Frequencies in the range 8–12 mc/sec were used.

The three independent elastic constants were evaluated, a density of 2.331 (measured by pycrometer) being used. Data and formulas used are summarized in Table I. Two crystals were measured—as indicated—with data obtained from the larger one being used to determine the elastic constants. Check measurements were made for the smaller crystal; and despite the less accurate "pulse overlap" technique used for two of the measurements, velocity agreement to within 0.15 per cent was obtained.

Both crystals were of a high degree of crystalline perfection as shown by etching and X-ray tests.

Domain Wall Relaxation in Nickel. W. P. MASON[1]. Letter to the Editor. *Phys. Rev.*, **83**, pp. 683–684, Aug. 1, 1951.

Phase Transitions in Ferroelectrics. B. MATTHIAS[1]. National Research Council, Comm. on Solids. *Phase Transformations in Solids.* Ed. by R. Smoluchowski, J. E. Mayer, W. A. Weyl. N. Y., Wiley, 1951. 660 pp.

[1] Bell Tel. Labs.
[5] Ill. Bell Tel. Co

Under the name ferroelectrics are classified those materials which exhibit dielectric anomalies phenomenologically similar to the magnetic behavior of the ferromagnetics. Perhaps it would have been more logical to use the term Rochelle-electrics, thus emphasizing the similarity in the dielectric behavior to that of Rochelle salt.

In this paper the known ferroelectrics are listed first, and then there follows a discussion of the various theories which have been created to explain them.

*Data on Random-Noise Requirements for Theater Television.** P. MERTZ[1]. *Jl. S.M.P.T.E.*, **57**, pp. 89–107, Aug., 1951.

Provisional evaluation of permissible random noise for theater television is considered from several sources of information. These cover broadcast television experience and the graininess in motion picture film; the requirements deduced from the various sources generally agree. For broadcast television, a frequency weighting and limit on weighted noise power have been used. The finer picture detail of theater television presumes a lower permissible random noise. Changes in weighting curve are discussed. A limit figure of noise is suggested, which is comparable to graininess effects in motion pictures, though slightly more severe than present published performance on camera tubes.

*A Spatial Harmonic Traveling-Wave Amplifier for Six Millimeters Wavelength.** S. MILLMAN[1]. *Proc. I.R.E.*, **39**, pp. 1035–1043, Sept., 1951.

This paper describes a traveling-wave amplifier in which the electron beam interacts with a spatial harmonic of an electromagnetic wave propagating along an array of resonator slots. The result is a considerable reduction in operating beam voltage for a given physical separation of the circuit elements. This type of amplifier operating at about 1,200 volts has yielded net power gains of about 18 db in the 6-mm wavelength region. A magnetic field of about 1,600 gauss is sufficient for proper beam focusing. Aside from small variations of gain with frequency that is caused by internal reflections, the bandwidth is of the order of 3 per cent.

*Form of Transient Currents in Townsend Discharges with Metastables.** J. P. MOLNAR[1]. *Phys. Rev.*, **83**, pp. 933–940, Sept. 1, 1951.

The form of the current is calculated for a Townsend discharge stimulated by a pulsed light beam, with particular reference to the current component initiated by metastable effects. The calculation is directed particularly to the development of methods for quantitative interpretation of current patterns observed experimentally.

*Studies of γ-Processes of Electron Emission Employing Pulsed Townsend Discharges on a Millisecond Time Scale.** J. P. MOLNAR[1]. *Phys. Rev.*, **83**, pp. 940–952, Sept. 1, 1951.

* A reprint of this article may be obtained on request.
[1] Bell Tel. Labs.

The relative amounts of electron emission from the cathode in a Townsend discharge caused by ions, photons, and metastables have been studied experimentally for several cathodes in argon, using pulsed-light stimulation of the discharge. The current initiated by metastables exhibits a slow build-up and decay, thus permitting easy separation from the faster rising effects of gas ionization and electron emission by photons and ions. Time constant studies of the slow component yielded a diffusion constant for metastable argon atoms of 45 $cm^2\ sec^{-1}$ at one millimeter pressure. The efficiencies of electron emission by metastables and ions was found to be closely the same, while the quantum yield for photon emission was found to be generally smaller.

*Electrical Properties of αFe_2O_3 and αFe_2O_3 Containing Titanium.** F. J. MORIN[1]. *Phys. Rev.*, **83**, pp. 1005–1010, Sept. 1, 1951.

Electrical conductivity, Hall effect, and Seebeck effect have been measured on two sets of polycrystalline samples of αFe_2O_3 and αFe_2O_3 containing from 0.05 to 1.0 atomic per cent titanium (n-type impurity). One set of samples contained 0.6 atomic per cent excess of iron (n-type impurity), the second set contained 0.6 atomic per cent deficiency of iron (p-type impurity).

The conductivity of pure αFe_2O_3 is independent of this amount of stoichiometric deviation. The slope of the log conductivity vs reciprocal temperature plot is 1.17 ev and the intercept at $1/T = 0$ is 2.1×10^4 $ohm^{-1}\ cm^{-1}$. Room temperature conductivity varies from -10^{-14} $ohm^{-1}\ cm^{-1}$ (extrapolated) for pure αFe_2O_3 to 0.3 $ohm^{-1}\ cm^{-1}$ for αFe_2O_3 containing 1.0 atomic per cent titanium.

The measured Hall voltages seem to result entirely from magnetization of the samples, which are weakly ferromagnetic, and disappear above the ferromagnetic Curie temperature.

The temperature variations of the Fermi level are determined from Seebeck data. The temperature variations of carrier concentration are determined from Fermi level and of mobility from carrier concentration and conductivity for some samples. Carrier concentration results indicate that each added titanium ion donates approximately one electron to the conduction process. Mobilities are found to be less than 2.0 cm^2/volt sec, suggesting that conduction involves electrons in the d level of iron.

*Acceptance Inspection of Purchased Material.** J. E. PALMER[3] and E. G. D. PATERSON[1]. *Ind. Quality Control*, **8**, pp. 15–19, Sept., 1951.

This paper describes some of the principles and procedures employed in the inspection of purchased material in the form of components or finished products. The authors' experience has been largely with procedures used in the Bell System, and the illustrations have therefore been drawn from this source. It is felt, however, that considerations leading to the choice of specific inspection tech-

* A reprint of this article may be obtained on request.
[1] Bell Tel. Labs.
[3] W. E. Co.

niques will be generally applicable even though the number and volume of items purchased and the number of suppliers involved may in some cases differ widely. In this presentation, stress has been placed on a discussion of the broader gauge factors underlying the engineering planning of inspection procedures rather than on specific sampling and control techniques.

*Analysis of Audio-Frequency Atmospherics.** R. K. POTTER[1]. *Proc I.R.E.*, **39**, pp. 1067–1069, Sept., 1951.

Sound portrayal techniques used in studies of speech and noise reveal the structure of atmospheric disturbances well known to long-wave radio and ocean-cable engineers as "whistlers," "swishes," and tweeks." It is suggested that renewed investigation of these effects, using modern analyzing tools, might yield information of considerable scientific interest.

*Reflection of Electromagnetic Waves from Slightly Rough Surfaces.** S. O. RICE[1]. *Communications on Pure and Applied Math.*, **4**, pp. 351–378, Aug., 1951.

*Color Television and Colorimetry.** W. T. WINTRINGHAM[1]. *Proc. I.R.E.*, **39**, pp. 1135–1172, Oct., 1951.

The high lights of the history of color measurement and of color photography are reviewed. Following this introduction, the principles of modern three-color colorimetry are developed from a hypothetical experiment in color matching. The conventional theory of "perfect color reproduction" by color television is built up from colorimetric background. Some of the difficulties to be expected in applying colorimetry to color television are brought out.

Finally, there is some discussion which tends to show that colorimetry may not be a sufficiently powerful tool to provide answers to all of the questions which will arise in the reproduction of scenes in color by television. The advantage of colorimetry as a background is indicated, however.

* A reprint of this article may be obtained upon request.
[1] Bell Tel. Labs.

Contributors to this Issue

CHARLES CLOS, C.E., New York University, 1927; New York Telephone Company, plant extension engineering, valuation and depreciation matters, intercompany settlements and tandem and toll fundamental plans, 1927–47. Pratt Institute, Evening School, Mathematics Instructor, 1946–49. Bell Telephone Laboratories, studies on development planning for local and toll switching systems and research in switching probability, 1947–. Member of A.I.E.E., New York Electrical Society, Mathematical Association of America, A.A.A.S., American Statistical Association, Iota Alpha, and Tau Beta Pi.

A. B. CRAWFORD, B.S. in E.E., Ohio State University, 1928; Bell Telephone Laboratories, 1928–. As a member of the Radio Research Department, he has been concerned with ultra short wave apparatus, measuring techniques, and propagation, and with microwave apparatus, measuring techniques, and propagation, as well as microwave radar and microwave antenna research. Member of I.R.E., Sigma Xi, Tau Beta Pi, Eta Kappa Nu, and Pi Mu Epsilon.

O. E. DE LANGE, B.S., University of Utah, 1930; M.A., Columbia University, 1937. Bell Telephone Laboratories, 1930–. Mr. De Lange has been engaged in radio research, including studies on high-frequency transmitters and receivers, frequency modulation, radar, broad-band systems, and pulse systems. Associate member of the I.R.E.

C. L. HOGAN, B.S. in Ch.E., Montana State College, 1942; M.S. in Physics, Lehigh University, 1947; Ph.D. in Physics, Lehigh, 1950. Anaconda Copper Mining Co., Great Falls, Montana, 1942–43. U. S. Navy, 1943–46. Instructor in Physics, Lehigh, 1947–50. Bell Telephone Laboratories, 1950–. Dr. Hogan has engaged in development work on boro-carbon resistors and microwave gyrators. Gold medal award for "Outstanding Engineer in Graduating Class," Montana State College, 1942. Letter of Merit from Chief of Naval Operations for work done in establishing and maintaining the acoustical torpedo shop at Pearl Harbor, 1944–46. Member of American Physical Society, Sigma Xi, Tau Beta Pi, and Phi Kappa Phi.

W. C. JAKES, JR., B.S.E.E., Northwestern University, 1944; M.S., Northwestern University, 1947; Ph.D., Northwestern University, 1949;

U.S. Navy, Airborne Radar Maintenance, 1944–46; Bell Telephone Laboratories, 1949–. Dr. Jakes has been engaged in microwave antenna and propagation studies. Member of I.R.E., Sigma Xi, Pi Mu Epsilon, and Eta Kappa Nu.

W. P. MASON, B.S. in E.E., University of Kansas, 1921; M.A., Ph.D., Columbia, 1928. Bell Telephone Laboratories, 1921–. Dr. Mason has been engaged principally in investigating the properties and applications of piezoelectric crystals, in the study of ultrasonics, and in mechanics. Fellow of the American Physical Society, Acoustical Society of America and Institute of Radio Engineers and member of Sigma Xi and Tau Beta Pi.

H. J. MCSKIMIN, B.S., University of Illinois, 1937; M.S., New York University, 1940. Bell Telephone Laboratories, 1937–. Here he has worked chiefly on crystal filters, piezoelectric elements, ADP crystals, studies of the acoustic properties of liquids and solids. Fellow of Acoustical Society of America and Member of Eta Kappa Nu and Sigma Xi.

R. S. OHL, B.S. in Electro-Chemical Engineering, Pennsylvania State College, 1918; U. S. Army, 1918 (2nd Lieutenant, Signal Corps); Vacuum tube development, Westinghouse Lamp Company, 1919–21; Instructor in Physics, University of Colorado, 1921–22. Department of Development and Research, American Telephone and Telegraph Company, 1922–27; Bell Telephone Laboratories, 1927–. Mr. Ohl has been engaged in various exploratory phases of radio research, the results of which have led to numerous patents. For the past ten or more years he has been working on some of the problems encountered in the use of millimeter radio waves. Member of American Physical Society and Alpha Chi Sigma and Senior Member of the I.R.E.

E. E. SUMNER, B.M.E., Cooper Union, 1948, holding Schweinburg Scholarship throughout entire college curriculum; Instructor of Physics, Cooper Union, 1947–48; Non-resident instructor of Massachusetts Institute of Technology, *Probability and Statistics—Applications to Sampling and Quality Control*, summer, 1950; Bell Telephone Laboratories, 1948–. Mr. Sumner was given rotational assignments in apparatus, switching, and television transmission development and switching research, and has worked on a number of projects, including the card translator, the magnetic drum, the video transmission evaluator, and the vibrating reed selector. He is currently engaged in the development of wire-spring relay. Member of Tau Beta Pi and Pi Tau Sigma.

ROGER I. WILKINSON, B.S. in E.E., 1924, Professional Engineer (honorary), 1950, Iowa State College, 1924; Northwestern Bell Telephone Company, 1920–21; American Telephone and Telegraph Company, 1924–34; Bell Telephone Laboratories, 1934–43 and 1946–. U. S. War Department, Washington and South Pacific, 1943–45. Mr. Wilkinson has been engaged in applications of the mathematical theory of probability to telephone problems. Medal for Merit, 1946. Member of A.S.E.E.; A.S.A.; Institute of Mathematical Statistics; American Society for Quality Control; Associate Member of A.I.E.E.; and Member of Eta Kappa Nu; Tau Beta Pi; Phi Kappa Phi; and Pi Mu Epsilon.