The Bell System Technical Journal

Vol. XVIII *July, 1939* *No. 3*

Frequency-Modulation: Theory of the Feedback Receiving Circuit

By JOHN R. CARSON

THIS paper may be regarded both as a continuation of a prior one by the writer and Thornton C. Fry [1] and as a companion of that by J. G. Chaffee [2] the inventor of the circuit under consideration. For an understanding of the present, an acquaintance with the prior paper [1] is absolutely necessary, since the fundamental analysis and the formulas there developed are too lengthy to be repeated here. References to that paper will be designated by (Ref.).

As the name implies, in the feedback circuit part of the incoming signal, after passing through a band-pass filter, a frequency detector [3] and a demodulator, is fed back through a variable frequency oscillator. The output of the variable frequency oscillator is connected to one branch of a modulator on the other branch of which the incoming high-frequency wave is impressed. While this method of feedback differs in some respects from that of the well known feedback amplifier, it is a fair inference that some if not all of the very important advantages of the feedback amplifier may also be present in the circuit under discussion. This inference is verified by the mathematical analysis of this paper.

After a brief development of the elementary theory and formulas of the feedback circuit as a receiver of frequency-modulated waves, the greater part of the paper is devoted to deriving formulas for the signal-to-noise power ratio—a criterion of fundamental importance in estimating the merits of the system. These are then compared with the corre-

[1] "Variable Frequency Electric Circuit Theory," *this Journal*, October 1937.

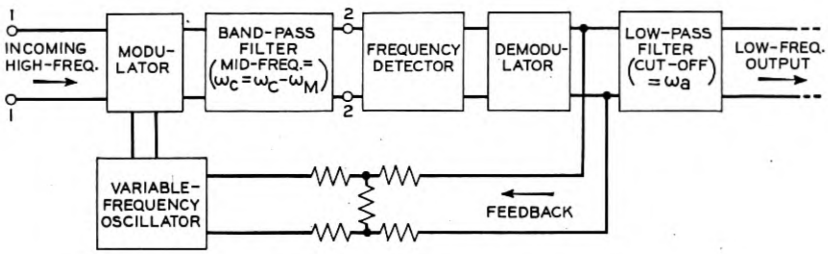
[2] "The Application of Negative Feedback to Frequency-Modulation Systems," *I. R. E. Proceedings*, May 1939; this issue of the *Bell Sys. Tech. Journal*.

[3] The function of the "frequency detector" is to detect or render explicit the variation of the "instantaneous frequency" of the frequency-modulated wave. A more precise term, therefore, would be "frequency variation detector," but for brevity the term used in the text is preferable.

sponding ratios formulated in (Ref.) for straight reception and also reception with amplitude limitation. In this way, as regards reduction of noise and "fading," the feedback circuit is found to have advantages comparable with those attainable by amplitude limitation.[4]

I

The receiving system operates as follows (see sketch):



Feedback receiving circuit.

The incoming frequency-modulated wave at terminals 1, 1 is taken as

$$E \exp\left(i\omega_C t + i\lambda \int^t s\,dt\right), \tag{1}$$

where E is the wave amplitude, ω_C the carrier frequency, λ the modulation index and $s = s(t)$ is the low-frequency signal which it is desired to recover.

This wave is impressed at terminals 1, 1 on one pair of terminals of a "product" modulator; on the other pair of terminals of the modulator there is impressed the output of a local variable-frequency oscillator:

$$M \exp\left(i\omega_M t + i\mu \int^t \sigma\,dt\right). \tag{2}$$

Here ω_M is the "carrier" frequency of the oscillator, μ (a positive real quantity) is the index of modulation of the oscillator and $\sigma = \sigma(t)$ is the low-frequency current fed back to the oscillator.

The output wave of the modulator is then equal to

$$c_1 EM \exp\left(i(\omega_C - \omega_M)t + i\int^t (\lambda s - \mu\sigma)dt\right)$$
$$+ c_1 EM \exp\left(i(\omega_C + \omega_M)t + i\int^t (\lambda s + \mu\sigma)dt\right). \tag{3}$$

[4] Armstrong, *Proc. I. R. E.*, May 1936, also see (Ref.).

The second term of (3) is suppressed by the band-pass filter.[5] Then writing $\omega_C - \omega_M = \omega_c$, it follows that the effective output wave is

$$c_1 EM \exp\left(i\omega_c t + i \int^t (\lambda s - \mu\sigma) dt \right). \tag{4}$$

ω_c is the intermediate carrier frequency and is always $< \omega_C$, the transmitted carrier frequency. The constant c_1 is a parameter depending on the characteristics of the modulator.

The wave (4) is transmitted through the band-pass filter, and the wave arriving at terminals 2, 2 is then

$$c_1 c_2 EM \exp\left(i\omega_c t + i \int^t (\lambda s - \mu\sigma) dt \right). \tag{5}$$

The parameter c_2 (taken as a constant) depends on the transmission characteristics from the modulator to terminals 2, 2.

Assuming an ideal frequency detector (see Ref.) the output to the terminals of the rectifier (or demodulator) is

$$c_1 c_2 c_3 EM \left(1 + \frac{\lambda s - \mu\sigma}{\omega_1} \right) \exp\left(i\omega_c t + i \int^t (\lambda s - \mu\sigma) dt \right). \tag{6}$$

Here the parameters c_3 and ω_1 depend on the characteristics of the frequency detector.

Finally assuming that

$$\frac{\lambda s - \mu\sigma}{\omega_1} < 1 \tag{7}$$

the low-frequency output of the rectifier[6] is

$$c_1 c_2 c_3 c_4 EM \left(1 + \frac{\lambda s - \mu\sigma}{\omega_1} \right). \tag{8}$$

If the constant term of (8) is suppressed and a fraction η of the rectified output is fed back to the oscillator we have, finally,

$$\sigma = \frac{m}{\mu} \frac{\lambda s}{1 + m}, \tag{9}$$

[5] Indeed the principal function of the band-pass filter is to suppress frequencies in the neighborhood of $\omega_C + \omega_M$.

[6] More generally a demodulator. In the present paper a straight-line rectifier is postulated for mathematical simplicity but the theory applies equally well to other forms of detection or demodulation.

where

$$m = c_1 c_2 c_3 c_4 \frac{\eta \mu E M}{\omega_1}, \\ = C \frac{\eta \mu E M}{\omega_1}. \tag{10}$$

The low-frequency current delivered to the receiver through the low-frequency output circuit proper differs from σ as given by (9) by a constant factor only.

From the foregoing we note that

$$\lambda s - \mu \sigma = \frac{\lambda s}{1 + m} \tag{11}$$

and that the "instantaneous frequency" of the intermediate high-frequency wave (4) is

$$\omega_c + \frac{\lambda s}{1 + m}. \tag{12}$$

Hereinafter, without any loss of generality, we suppose that $-1 \leqq s(t) \leqq 1$. Consequently the intermediate frequency-modulated wave has a frequency variation lying between $\pm \lambda/(1 + m)$, whereas in the incoming frequency-modulated wave, the frequency variation lies between $\pm \lambda$.

We note also from (9) that if the parameter m is large compared with unity, the low-frequency received wave is approximately given by

$$\sigma = \frac{\lambda}{\mu} s. \tag{13}$$

The recovered signal is thus (for large values of m) seen to be independent of the amplitude, E, of the incoming high-frequency wave; therefore, the system is insensitive to "fading."

II

We now take up the problem of calculating the relative low-frequency *noise* and *signal* powers, the ratio of which is of fundamental importance in appraising the merits of the receiving circuit. In this we shall closely follow the methods developed in Section IV (Ref.).

We suppose that at terminals 1, 1 there enters, in addition to the signal, a typical noise element

$$a_n \exp\left(i(\omega_C + \omega_n)t + i\alpha_n\right). \tag{14}$$

We write for convenience in the subsequent analysis

$$a_n = A_n E. \tag{15}$$

A_n is then the *relative* amplitude of the noise element, referred to the amplitude E of the high-frequency signal. We shall suppose throughout that A_n is small compared with unity; that is, the noise is small compared with the signal.

We further suppose that at terminals 2, 2 between the band-pass filter and the frequency detector there is introduced a second typical noise element

$$b_n \exp\,(i(\omega_c + \omega_n)t + i\beta_n), \tag{16}$$

which is entirely independent of the noise element (14). This may be regarded as caused by tube-noise in amplifiers (not shown in sketch). We write

$$b_n = B_n E \tag{17}$$

so that B_n is the relative amplitude of the noise element, referred to the amplitude E of the incoming signal wave. It also is assumed small compared with unity.

The total input to the frequency detector, neglecting the random phase angles, is then

$$c_1c_2EM\left[\exp\left(i\int^t \Omega\,dt\right) + A_n\exp\left(i\int^t(\Omega + \Omega_n{}^a)dt\right) + \frac{B_n}{c_1c_2M}\exp\left(i\int^t(\Omega + \Omega_n{}^b)dt\right)\right], \tag{18}$$

where

$$\begin{aligned} \Omega &= \omega_c + \lambda s - \mu\sigma \\ \Omega_n{}^a &= \omega_n - \lambda s \\ \Omega_n{}^b &= \omega_n - \lambda s + \mu\sigma. \end{aligned} \tag{19}$$

The output of the frequency detector is then (see Ref.)

$$\begin{aligned} c_1c_2c_3EM\exp\left(i\int^t \Omega\,dt\right) & \\ \times\Bigg[1 + \frac{1}{\omega_1}(\lambda s - \mu\sigma) & \\ + A_n\left(1 + \frac{1}{\omega_1}(\omega_n - \mu\sigma)\exp\left(i\int^t \Omega_n{}^a\,dt\right)\right) & \\ + \frac{B_n}{c_1c_2M}\left(1 + \frac{\omega_n}{\omega_1}\right)\exp\left(i\int^t \Omega_n{}^b\,dt\right)\Bigg]. & \end{aligned} \tag{20}$$

After the output as given by (20) is rectified, the constant term suppressed, and only first powers in A_n and B_n retained, we get finally

$$\sigma = \left(\frac{m/\mu}{1+m}\right)\left[\left(\lambda s + A_n\left(\omega_1 + \omega_n - \frac{m}{1+m}\right)\lambda s\right)\cos\int^t \Omega_n^a\,dt + \frac{B_n}{c_1c_2M}(\omega_1+\omega_n)\cos\int^t \Omega_n^b\,dt\right]. \quad (21)$$

Now the right-hand side of (21) corresponds precisely with formula (64) (Ref.) on which the calculation of the relative low-frequency noise and signal powers is based. Consequently following the methods developed in Ref. and assuming A_n and B_n small we get

$$\overline{\sigma^2} = \left(\frac{m/\mu}{1+m}\right)^2\left[\lambda^2\overline{s^2} + \left(\frac{1}{3}\omega_a^2 + \omega_1^2 + \frac{\lambda^2\overline{s^2}}{(1+m)^2}\right)\omega_a N_a^2 + \left(\frac{1}{3}\omega_a^2 + \omega_1^2 + \overline{(\lambda s - \mu\sigma)^2}\right)\omega_a N_b^2/c_1^2c_2^2M^2\right]. \quad (22)$$

The *relative* low-frequency noise and signal powers are then (omitting the common factor $\left(\frac{m/\mu}{1+m}\right)^2$)

$$P_N = \frac{1}{3}\omega_a^3N_a^2\left[1 + 3\left(\frac{\omega_1}{\omega_a}\right)^2 + 3\,\frac{(\lambda/\omega_a)^2}{(1+m)^2}\,\overline{s^2}\right] + \frac{1}{3}\,\frac{\omega_a^3N_b^2}{c_1^2c_2^2M^2}\left[1 + 3\left(\frac{\omega_1}{\omega_a}\right)^2 + 3\,\frac{\overline{(\lambda s - \mu\sigma)^2}}{\omega_a^2}\right], \quad (23)$$

$$P_S = \lambda^2\overline{s^2}.$$

In these formulas N_a^2 is proportional to the noise power level in the neighborhood of the carrier frequency ω_C; it enters at the input terminals 1, 1 (see Ref. Appendix 2). N_b^2 is proportional to the noise power level in the neighborhood of the intermediate carrier frequency ω_c; it enters at terminals 2, 2. ω_a is the highest essential frequency in the low-frequency signal $s(t)$; it is the cut-off frequency of the low-pass filter.

Formula (22) is solvable (see Appendix) but a simple approximate solution, valid when the noise is small compared with the signal, is made possible by observing that under this restriction

$$\lambda s - \mu\sigma = \frac{\lambda s}{1+m} \quad (11)$$

to a good approximation. Introducing this approximation into P_N as given by (23) and writing

$$N^2 = N_a^2 + N_b^2/c_1^2 c_2^2 M^2, \tag{24}$$

we have

$$P_N = \frac{1}{3}\,\omega_a^3 N^2 \left(1 + 3\left(\frac{\omega_1}{\omega_a}\right)^2 + 3\,\frac{(\lambda/\omega_a)^2}{(1+m)^2}\,\overline{s^2} \right), \tag{25}$$

$$P_S = \lambda^2 \overline{s^2}.$$

Now from the inequality, necessary for rectification,

$$\omega_1 > \frac{\lambda}{1+m}$$

it is seen that as the parameter m is increased, ω_1 may be reduced by the factor $1/(1 + m)$. In accordance with this, we replace ω_1 by $\omega_1/(1 + m)$ in (25) and get

$$P_N = \frac{1}{3}\,\omega_a^3 N^2 \left[1 + \frac{3}{(1+m)^2}\left(\left(\frac{\omega_1}{\omega_a}\right)^2 + \left(\frac{\lambda}{\omega_a}\right)^2 \overline{s^2}\right)\right], \tag{26}$$

$$P_S = \lambda^2 \overline{s^2}.$$

The noise power P_N can be still further reduced by eliminating ω_1 from (26) by a circuit arrangement explained in Ref. Section III; if this is done we get, instead of (26),

$$P_N = \frac{1}{3}\,\omega_a^3 N^2 \left(1 + 3\,\frac{(\lambda/\omega_a)^2}{(1+m)^2}\,\overline{s^2} \right),\qquad P_S = \lambda^2 \overline{s^2}. \tag{27}$$

We have now to compare the relative noise and signal powers of the feedback with (1) straight reception *without* feedback and (2) reception with *amplitude limitation.*

For *straight reception* (without feedback) we have (see equation (68), Ref.), corresponding to (26),

$$P_N = \frac{1}{3}\,\omega_a^3 N^2 \left(1 + 3\left(\frac{\omega_1}{\omega_a}\right)^2 + 3\left(\frac{\lambda}{\omega_a}\right)^2 \overline{s^2} \right),\qquad P_S = \lambda^2 \overline{s^2}, \tag{28}$$

and corresponding to (27)

$$P_N = \frac{1}{3}\,\omega_a^3 N^2 \left(1 + 3\left(\frac{\lambda}{\omega_a}\right)^2 \overline{s^2} \right),\qquad P_S = \lambda^2 \overline{s^2}. \tag{29}$$

When noise reduction is effected by *amplitude limitation* the corresponding relative noise and signal powers (see equation (78), Ref.) are

$$P_N = \frac{1}{3}\omega_a^3 N^2, \qquad (30)$$

$$P_S = \lambda^2 \overline{s^2}.$$

If we assume as above that $-1 \leq s \leq 1$ and $\overline{s^2}$ is of the order of magnitude of 1/2, then in practical applications $\lambda/\omega_a \gg 1$ and $\omega_1 > \lambda$. On this basis comparison of (26) with (28) and (27) with (29) shows that, when $m \gg 1$, the noise power with *feedback* is very much smaller than *without feedback*, the ratio of the noise powers in the two cases being approximately $1/(1+m)^2$. (This assumes, of course, that N^2 is approximately equal in the two cases.)

Comparing, however, the noise power with *feedback* to that obtainable by *amplitude limitation*, it will be seen that in order to reduce the former to the order of magnitude of the latter it is necessary that

$$\frac{(\lambda/\omega_a)}{(1+m)} < 1. \qquad (31)$$

From the preceding it is seen that the performance of the feedback circuit and the reduction in noise-power ratio obtainable depend in a fundamental manner on the parameter m, defined above by the formula

$$m = c_1 c_2 c_3 c_4 \frac{\eta \mu E M}{\omega_1}. \qquad (32)$$

If the characteristics of the modulator rectifier and variable-frequency oscillator are stipulated, it is possible to calculate m in terms of these characteristics and the constants and connections of the network. It is experimentally determinable (among other ways) as follows:

Let the feedback circuit be opened between the low-pass filter and the variable-frequency oscillator, and let the filter be closed by an impedance equal to that of the oscillator as seen from the filter. Then $m = 0$ (since there is no low-frequency feedback to the oscillator) but m/μ is finite.

Denoting the value of σ under these circumstances by σ_1, it follows from (9) that

$$\sigma_1 = \frac{m}{\mu}\lambda s. \qquad (33)$$

Consequently dividing σ_1 by σ, as given by (9), we have

$$1 + m = \sigma_1/\sigma,$$

$$m = \frac{\sigma_1 - \sigma}{\sigma}. \qquad (34)$$

Stated in words, $1 + m$ is the reciprocal of the ratio of the values of σ *without* and *with* the low-frequency feedback into the oscillator. It should be noted that this requires that the band-pass filter transmit the frequency band 2λ centered on ω_c.

APPENDIX

In formula (22), the expression $\overline{(\lambda s - \mu\sigma)^2}$ has been replaced by $\lambda^2\overline{s^2}/(1 + m)^2$, its value when the noise is absent. When noise is present, but small compared with the signal, this should still give a good approximation for $\overline{\sigma^2}$. We now propose to derive an exact solution of (22); to this end we write

$$\lambda s - \mu\sigma = \frac{\lambda s}{1 + m} - n(t) \tag{1a}$$

which is always possible.

Now inspection of (1a) shows that $n(t)$ is the value of $\mu\sigma$ when $s = 0$; consequently $\overline{n^2} = \mu^2\overline{\sigma_0^2}$ where $\overline{\sigma_0^2}$ is given by (34). Furthermore, since s and n are entirely independent, $\overline{sn} = 0$, and

$$\overline{(\lambda s - \mu\sigma)^2} = \frac{\lambda^2\overline{s^2}}{(1 + m)^2} + \mu^2\overline{\sigma_0^2}\,. \tag{2a}$$

Substitution of (2a) in (22) gives for P_N, instead of (23),

$$P_N = \frac{1}{3}\,\omega_a^3 N^2\left(1 + 3\left(\frac{\omega_1}{\omega_2}\right)^2 + 3\,\frac{(\lambda/\omega_a)^2}{(1 + m)^2}\,\overline{s^2}\right) + \omega_a\mu^2\sigma_0^2 N_b^2/c_1^2c_2^2M^2. \tag{3a}$$

The second term is a second order quantity in the noise power and may therefore be neglected when the noise is small, as is assumed throughout this paper.

The Application of Negative Feedback to Frequency-Modulation Systems *

By J. G. CHAFFEE

Negative feedback can be applied to a frequency-modulation receiver of superheterodyne type by causing a portion of the output voltage to frequency-modulate the local oscillator in such phase as to reduce the output signal. As a consequence of this arrangement the effective frequency modulation of the intermediate wave is diminished by the feedback factor. This reduction is accompanied by a decrease in noise and distortion. Restoration of the original signal level by increasing the degree of modulation at the transmitter brings about a corresponding increase in signal-to-noise ratio provided the disturbance is not too great, while distortion ratios are improved to about the same extent. These effects are treated analytically for the case where the disturbance level is sufficiently low to permit simplifying assumptions to be made. The results are in general agreement with observations made on an experimental laboratory system.

Comparing the feedback system with a frequency-modulation system using amplitude limitation, the ratio of signal level to noise level in the absence of modulation is identical in two systems. During modulation the noise level increases in the feedback system by an amount depending upon the ratio of the effective frequency shift of the intermediate-frequency wave to the signal band width. By keeping this ratio small, the increase in noise during modulation can be made relatively unimportant.

In cases where the disturbance level is high, phenomena have been observed which are very similar to those encountered when amplitude limitation is used.

Introduction

THIS paper describes a method for improving the performance of receivers designed to receive frequency-modulated waves. In its broader aspects this method can be described as the application of the principle of negative feedback to a superheterodyne frequency-modulation receiver. In its details the application of the feedback principle necessitates the use of a rather unusual circuit arrangement. This circuit differs from that of the simple feedback system in that the voltages fed back are not of the same frequency as those applied

* Presented before New York Section of I. R. E., May 3, 1939. Published in *Proceedings, I. R. E.*, May 1939.

to the input of the receiver, and are caused to influence the response of the system by modifying the performance of the modulator.

In the ordinary feedback amplifier a part of the output voltage is carried back to the input and there combined with the applied voltage. The result is to modify the output and if the gain of the system is thereby reduced the feedback is said to be negative. The many advantages which result from negative feedback have been described by Black [1] and are coming to be more generally appreciated. The present paper deals with a method for adapting this principle to a frequency-modulation receiver and will show an example of its application to an experimental system in the laboratory.

General Discussion

Method of Applying Feedback

Consider a frequency-modulation receiver in which the incoming wave is combined with the output of a local oscillator in a modulator to produce a wave of intermediate frequency. This is then amplified, converted into an amplitude-modulated wave, and finally detected. The frequency of the intermediate wave is equal to the instantaneous difference in the frequencies of the incoming carrier and the local oscillator. So long as the frequency of this oscillator remains fixed the intermediate wave will be frequency-modulated in exact correspondence with the incoming wave. Suppose now that the local oscillator is frequency-modulated from a source of the same frequency and phase as that applied to the transmitter. As the index of modulation at the local oscillator is increased from zero the extent to which the intermediate wave is modulated will diminish since its instantaneous frequency is equal to the difference in the frequencies of the two sources. It then follows that if these two devices are modulated to the same extent the difference frequency will become constant and the output of the system will be zero. Finally a further increase in modulation of the local oscillator will cause the intermediate wave to be modulated with a 180-degree phase reversal.

This process can be readily analyzed as follows: Assume the oscillator at the transmitter to have been frequency-modulated by the signal wave

$$e = E_1 \cos pt. \tag{1}$$

The voltage delivered to the modulator by the incoming wave will be

$$A \cos (\omega_1 t + x_1 \sin pt + \phi_1) \tag{2}$$

[1] H. S. Black, "Stabilized Feedback Amplifiers," *Elec. Engg.*, vol. 53, pp. 114–120, January 1934.

where $x_1 = \Delta\omega_1 \div p = \rho E_1 \div p$, and $\Delta\omega_1$ is 2π times the maximum frequency shift. The local oscillation impresses the wave

$$B \cos (\omega_2 t + x_2 \sin pt + \phi_2) \tag{3}$$

where

$$x_2 = \Delta\omega_2 \div p.$$

Application of these waves to a square-law modulator will yield a difference frequency wave proportional to

$$AB \cos [(\omega_1 - \omega_2)t + (x_1 - x_2) \sin pt + \phi_1 - \phi_2] \tag{4}$$

for the case where $\omega_1 > \omega_2$, or when the reverse is true

$$AB \cos [(\omega_2 - \omega_1)t - (x_1 - x_2) \sin pt + \phi_2 - \varphi_1]. \tag{5}$$

In either case the resultant modulation index of the intermediate wave is the numerical difference of the original indexes, the difference in sign

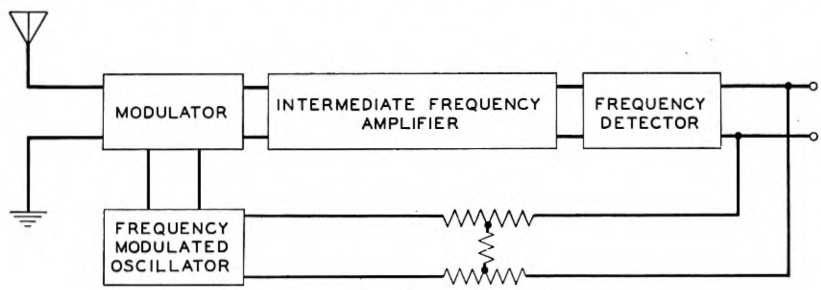


Fig. 1—Basic feedback circuit.

in the two cases signifying that the detected outputs will be of opposite phase. If $x_1 = x_2$ the modulation is reduced to zero, and if $x_2 > x_1$ modulation reappears with a phase reversal. It might be noted that if x_2 were originally made negative, thus causing the two oscillators to be frequency-modulated in opposite phase, the apparent modulation of the incoming wave could be increased indefinitely.

Suppose, now, that instead of frequency-modulating the local oscillator from an independent source, the equivalent is accomplished in a practical way. For this purpose a voltage from the output of the receiver is impressed upon the local oscillator as shown in Fig. 1. The transmitted wave will then have a modulation index $x_1 = \rho_1 E_1 \div p$, while the local oscillator, being acted upon by a portion of the output voltage E_0, will have an index $x_2 = k\rho_2 E_0 \div p$. If the frequency detector [2] is assumed to be linear the amplitude of the detected output

[2] The term *frequency detector* is used in this paper to designate the combination of conversion circuit and amplitude detector. A more extended discussion of modulation and detection is given in Appendix A.

will be proportional to the product of the amplitudes A and B of the incoming and local oscillator waves, the resultant index of the intermediate wave, and the slope factor a_1. Thus we can write the output voltage amplitude

$$E_0 = \alpha a_1 AB(x_1 - x_2)p = \alpha a_1 AB(\rho_1 E_1 - k\rho_2 E_0). \tag{6}$$

Therefore

$$E_0 = \frac{\alpha a_1 AB\rho_1 E_1}{1 + \alpha k a_1 AB\rho_2}. \tag{7}$$

Setting $\alpha a_1 AB = \mu$ and $k\rho_2 = -\beta$ we obtain the familiar form encountered in the analysis of feedback amplifiers

$$E_0 = \frac{\mu(\rho_1 E_1)}{1 - \mu\beta}. \tag{8}$$

Without feedback the output of the system is merely $\mu(\rho_1 E_1)$. The feedback factor $1 + \alpha a_1 kAB\rho_2$ is a measure of the extent to which the over-all gain of the system has been modified by feedback. If this factor is greater than unity the feedback is negative, while if k is made negative by reversing the feedback connections the effect is regenerative, and instability is encountered when the factor becomes zero.

It will be noted that when $\alpha a_1 kAB\rho_2 \gg 1$, (7) becomes

$$E_0 = \frac{\rho_1 E_1}{k\rho_2}. \tag{9}$$

Thus for large amounts of feedback, the output signal becomes independent of such factors as fading of the incoming wave, variations in the local oscillator voltage, or changes in detector efficiency. Hence automatic gain control is secured. This feature is equivalent to that found in ordinary feedback amplifiers in that for large amounts of feedback the over-all gain becomes independent of variations in the performance of the amplifier proper.

Reduction of Noise

The application of negative feedback in the manner described brings about a reduction in signal level by decreasing the effective modulation of the received wave. It then becomes possible to increase the modulation level at the transmitter to a corresponding degree and thus to restore the output signal to its former value. This process is made possible through the use of frequency rather than amplitude modulation since the permissible degree of modulation is then determined by the receiver characteristics. It will be shown that feedback also reduces the noise level at the output of the receiver, provided that the disturbance is not too great. Thus when the modulation level is raised to offset the effect of feedback an improvement in signal-to-noise ratio is realized.

The mechanism by which noise is reduced can be described qualitatively as follows: Noise at the output terminals of the receiver is caused to frequency-modulate the intermediate wave in such fashion as to produce, upon detection, a component which tends to cancel that which would exist in the absence of feedback. An analysis of this process for the case where the carrier is large compared with the disturbance responsible for the noise is developed [3] in Appendix B. It is assumed that the disturbance can be represented by a continuous spectrum of sinusoidal voltages of equal amplitude but phased at random. Impressed along with the disturbance is the signaling wave (2). Then if N^2 is the mean disturbing power per unit of band width in the vicinity of the carrier frequency, and r_1 is the resistance of the input circuit, it is shown that the output noise power is [4]

$$P_N = \frac{2N^2 r_1}{F^2}\left[a_0^2 + \frac{a_1^2\Delta\omega^2}{2F^2} + \frac{a_1^2 q_a^2}{3}\right] q_a \tag{10}$$

where a_0 and a_1 are, respectively, the gain and slope factor of the intermediate amplifier and conversion system as defined by (47), and q_a represents the upper limit of frequency response of the output circuit, or the upper limit of audibility as the case may be. F is the feedback factor $(1 - \mu\beta)$. The corresponding signal power is

$$P_s = \frac{A^2 a_1^2 \Delta\omega^2}{2F^2}. \tag{11}$$

The reduction in signal level occasioned by feedback can be offset by increasing the frequency shift of the transmitted wave. If it is increased so as to have the value $\Delta\Omega = F\Delta\omega$ then the shift of the inter-

[3] An analysis of the effect of feedback upon noise in this system was first developed by J. R. Carson by methods similar to those used in "Variable Frequency Electric Circuit Theory with Application to the Theory of Frequency Modulation," Carson and Fry, *Bell Sys. Tech. Jour.*, vol. 16, pp. 513–540, October 1937. This has been embodied in a paper by Mr. Carson entitled, "Frequency Modulation: Theory of the Feedback Receiving Circuit," published in this issue of the *Bell Sys. Tech. Jour.* Carson's treatment is more general in that an arbitrary signal wave is postulated whereas the analysis given in Appendix B is restricted to a sinusoidal signal wave. The methods used here are more elementary and may therefore appeal to a somewhat wider audience.

[4] The expressions for signal and noise power used in this section are relative. Factors determining their absolute magnitude are given in the Appendix. In all cases the symbol $\Delta\omega^2$ is to be taken as signifying $(\Delta\omega)^2$.

mediate-frequency wave will be restored to its original value $\Delta\omega$ and the signal level will remain unchanged. Then the noise power becomes

$$P_N = \frac{2N^2r_1}{F^2}\left[a_0{}^2 + \frac{a_1{}^2\Delta\Omega^2}{2F^2} + \frac{a_1{}^2q_a{}^2}{3} \right] q_a \tag{12}$$

which can be written

$$P_N = \frac{2N^2r_1}{F^2}\left[a_0{}^2 + \frac{a_1{}^2\Delta\omega^2}{2} + \frac{a_1{}^2q_a{}^2}{3} \right] q_a. \tag{12a}$$

The noise-to-signal power ratio is improved by the factor F^2, since

$$\frac{P_N}{P_s} = \frac{1}{F^2}\frac{4N^2r_1}{A^2}\left[\frac{a_0{}^2}{a_1{}^2\Delta\omega^2} + \frac{1}{2} + \frac{q_a{}^2}{3\Delta\omega^2} \right] q_a. \tag{13}$$

Of the factors in (12) the first is the result of modifications of the amplitude of the incoming wave by the disturbance. Although subject to reduction by feedback it can be balanced out completely by the use of differentially connected frequency detectors having slope factors a_1 and $-a_1$. The second term is dependent upon the degree of modulation of the intermediate wave. It is usually of less consequence in its effect upon the listener. The remaining term is the result of phase modulation of the signal wave by the disturbance. Under the condition that the output signal is held constant by increasing the transmitted band, all terms which contribute to the noise level in a given case are reduced alike by feedback.

If differential frequency-detection is employed (12a) becomes

$$P_N = \frac{2N^2r_1a_1{}^2}{F^2}\left[\frac{\Delta\omega^2}{2} + \frac{q_a{}^2}{3} \right] q_a. \tag{14}$$

During non-signaling periods the first term becomes zero. Hence during periods of modulation the background noise power is increased by the factor

$$1 + \frac{3}{2}\frac{\Delta\Omega^2}{F^2q_a{}^2} = 1 + \frac{3}{2}\frac{\Delta\omega^2}{q_a{}^2}\,. \tag{15}$$

If conditions are such that the maximum shift experienced by the intermediate-frequency wave is numerically equal to q_a, then the noise level will be increased by 4 decibels during periods of full modulation. In the experimental system to be described the ratio of $\Delta\omega$ to q_a was allowed to attain a value of 1.75, resulting in a maximum increase of 7.5 decibels.

In order to secure large noise reduction it is necessary to produce a frequency shift in the transmitted wave much greater than the signal band width. Thus, in common with frequency-modulation systems employing amplitude limiters,[5] this advantage is secured at the expense of band width. In this connection it is of interest to compare amplitude limitation and feedback systems on the basis of equal width of transmitted band. Hence it will be assumed that in each case the transmitted wave is modulated to the extent of $\pm\,\Delta\Omega = \pm\,F\Delta\omega$. In Fig. 2 are shown idealized characteristics of conversion systems which might be used in the two systems. The adjustment shown in Fig. 2(a) is suitable for use with the limiter system. With the feedback system that shown in Fig. 2(b) would be necessary to secure the same percentage of amplitude modulation after conversion. This represents

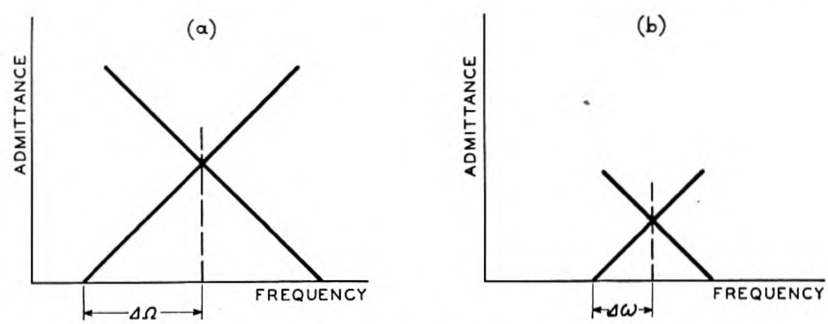


Fig. 2—Idealized conversion-system characteristics for (a) limiter system, (b) feedback system.

the minimum band width which could be provided in the conversion system with feedback, though several considerations make it desirable to use an adjustment lying somewhere between the two shown. The manner of tuning or the slope factors assumed in each case are immaterial to the present comparison provided that, in either system, the linear portion of the characteristic is of sufficient extent to effect proper conversion of the intermediate-frequency wave.

The noise-to-signal power ratio obtainable with the feedback system will be that given by (13) with the first term omitted since it is balanced out by the push-pull arrangement. Thus

$$\frac{P_N}{P_s} = \frac{4N^2r_1}{F^2A^2}\left[\frac{1}{2} + \frac{q_a^2}{3\Delta\omega^2}\right] q_a. \tag{16}$$

[5] E. H. Armstrong, "A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation," *Proc. I. R. E.*, vol. 24, pp. 689–740, May 1936.

In the system corresponding to Fig. 2(a), the signal power will be

$$\frac{A^2 a_1^2 \Delta\Omega^2}{2}. \tag{17}$$

Equation (10) can be used to determine the noise power level for a frequency-modulation system without amplitude limitation by setting $F = 1$. If balanced detection is used the term in a_0 becomes zero. It has been shown by Carson and Fry [3] that the addition of an ideal limiter removes all terms but the third, with either single or balanced detectors. Hence for the limiter system the noise ratio becomes

$$\frac{P_N}{P_s} = \frac{4N^2 r_1}{A^2}\left(\frac{q_a^3}{3\Delta\Omega^2}\right). \tag{18}$$

Since $\Delta\Omega = F\Delta\omega$ this can be put in a form similar to (16)

$$\frac{P_N}{P_s} = \frac{4N^2 r_1}{F^2 A^2}\left(\frac{q_a^2}{3\Delta\omega^2}\right) q_a. \tag{19}$$

Comparing (16) and (19) it is seen that the noise ratio in the feedback system is greater than that for the limiter system by the factor (15). This is a consequence of the increase in noise level which occurs during modulation in the former system. The ratio of noise level during non-signaling periods to signal level is identical in the two systems.

While the noise increment which appears during modulation is usually not of great consequence from a practical standpoint, it can be reduced by increasing the feedback factor beyond the point dictated by the signal band which it is permissible to transmit. In previous discussions it has been assumed that the application of a given amount of negative feedback is to be accompanied by a corresponding increase in modulation level at the transmitter. In this way the modulation of the intermediate frequency wave is kept constant so as to maintain a fixed signal level as the band width of the transmitted wave is increased. Having arrived at a limiting value of band spread the feedback factor can be increased further. Suppose that modulation of the transmitter is to be limited to a value of $\Delta\Omega = F_1\Delta\omega$, but that the feedback applied to the receiver is made to exceed F_1 by a factor which we shall call F_2. Then the actual feedback factor will be F_1F_2 and we have

$$P_s = \frac{A^2 a_1^2 \Delta\omega^2}{2F_2^2} \tag{20}$$

$$P_N = \frac{2N^2 r_1}{F_1^2 F_2^2}\left[\frac{a_1^2\Delta\omega^2}{2F_2^2} + \frac{a_1^2 q_a^2}{3}\right] q_a \tag{21}$$

giving

$$\frac{P_N}{P_s} = \frac{4N^2 r_1}{A^2 F_1^2}\left[\frac{1}{2F_2^2} + \frac{q_a^2}{3\Delta\omega^2}\right] q_a. \tag{22}$$

Thus the additional feedback represented by the factor F_2 is directly effective against the noise increment accompanying modulation. Reduction of this increment brings about a still closer correspondence between the limiter and feedback systems as is seen by setting $F = F_1$ in (19) and comparing with (22).

The above discussion and the analysis given in Appendix B are based upon the assumption that the carrier amplitude is large compared with that of the disturbance. A rigorous analysis, applicable to the case where this ratio is unrestricted, becomes exceedingly involved. However, a rough indication of what is to be expected in the presence of a high level of disturbance can be obtained quite simply from (52) developed in Appendix B. Assuming that modulation is not present this can be put in the simple form

$$\sigma = \frac{1}{F}\,\frac{Q'(a_0 + a_1\omega_n)}{1 + \dfrac{Q'}{A'}\left(\dfrac{F-1}{F}\right)\cos \omega_n t.}\cos \omega_n t. \tag{52a}$$

When $Q' \ll A'$ the wave form of the output noise produced by a single element of disturbance is very closely a sinusoid. However, when Q' and A' become comparable in magnitude the output wave becomes badly peaked when $\omega_n t = n\pi$. While the above expression is only a very rough approximation under these conditions, a plot of the wave form so obtained exhibits all of the essential characteristics of the curves given by Crosby in a recent paper [6] dealing with noise in frequency-modulation systems using amplitude limitation. These curves show a similar peaking of the output-noise wave form when the ratio of carrier to disturbance amplitude is in the vicinity of unity. The description given by Crosby of the manifestations of this phenomenon observed in an experimental system applies rather closely to what has been found in the feedback system. A more detailed account will be found in a later section.

Examination of (52a) shows that the output wave can assume very large and even infinite peak values when Q' and A' are approximately equal. The existence of high peak values of noise implies both a large instantaneous deviation in the frequency of the intermediate wave, and a conversion-circuit characteristic of unlimited extent. The finite

[6] Murray G. Crosby, "Frequency Modulation Noise Characteristics," *Proc. I. R. E.*, vol. 25, pp. 472–514, April 1937. The curves referred to are given in Fig. 4 of the above paper.

limits of the characteristic of the over-all intermediate-frequency system have the effect of holding the maximum peaks of noise to a value equal to the highest signal peaks obtainable in the absence of the disturbance. Furthermore, the existence of high noise peaks in the presence of modulation can result in the momentary assumption by the instantaneous intermediate frequency of values outside of the region to which the system is normally responsive. Thus the output signal will appear to be chopped by the higher noise peaks, and as a consequence its energy content will be considerably reduced.

The above effects are, of course, present in systems using limiters and have already been discussed in greater detail by Crosby.[6]

Distortion Reduction

One of the chief benefits which can be realized through the use of negative feedback is the reduction of non-linear distortion products generated in the forward branch of the system. While the distortion in properly designed amplifiers is sufficiently low for many purposes, cases frequently arise in which the requirements are much more severe. In an amplifier which is to handle several channels in a high grade multiplex system, the distortion products should be of the order of 60 decibels below the fundamental of the output. This degree of excellence is most readily obtained by using negative feedback.

In radio systems designed for multiplex service it is of equal importance that the distortion level be kept at a correspondingly low level if crosstalk is to be avoided. It is therefore of interest to inquire into the manner in which distortion is modified in the present feedback system.

An analysis of the effect of feedback upon distortion is given in Appendix A. If the transmitter is modulated with a signal wave $S = S(t)$ so that its instantaneous frequency becomes

$$\omega + \rho_1 S \tag{23}$$

then, in the presence of non-linearity in the receiver, the output of the system can be written as a power series in the variable frequency term $\rho_1 S$. Thus for the first three orders we shall have

$$\sigma = \alpha AB[b_1\rho_1 S + b_2(\rho_1 S)^2 + b_3(\rho_1 S)^3]. \tag{24}$$

If feedback is applied without altering the modulation level at the transmitter it is shown that the above series becomes

$$\sigma_F = \alpha AB\left(\frac{b_1}{F}\rho_1 S + \frac{b_2}{F^3}(\rho_1 S)^2 + \frac{1}{F^4}\left[b_3 - \frac{2b_2^2}{b_1}\left(\frac{F-1}{F}\right)\right](\rho_1 S)^3\right). \tag{25}$$

When the feedback factor F is large this can be written

$$\sigma_F = \alpha AB\left(\frac{b_1}{F}\rho_1 S + \frac{b_2}{F^3}(\rho_1 S)^2 + \frac{1}{F^4}\left[b_3 - \frac{2b_2^2}{b_1}\right](\rho_1 S)^3\right). \quad (26)$$

Upon increasing the modulation by the factor F so as to restore the original level of the fundamental, the output becomes

$$\sigma_F' = \alpha AB\left(b_1\rho_1 S + \frac{1}{F}\left[b_2(\rho_1 S)^2 + \left(b_3 - \frac{2b_2^2}{b_1}\right)(\rho_1 S)^3\right]\right). \quad (27)$$

Second order distortion products are reduced with respect to the fundamental level by the feedback factor. Third (and higher) order products are modified to an extent depending upon the relative values of the distortion coefficients and the amount of feedback. If, as can readily be the case when a balanced detecting system is used,

$$b_3 \gg \frac{2b_2^2}{b_1} \quad (28)$$

third order products are reduced in the same manner as those of second order. In any case, by applying sufficient feedback a point will be reached where a given increment in feedback will produce a corresponding reduction in all distortion products.

Equation (25) shows that the greatest improvement in distortion is obtained if the modulation level is not increased when feedback is applied. The large reductions result partly from feedback and in part from the fact that the system is operating at reduced percentage of modulation. Under this condition there is no improvement in background noise ratio, though the noise increment which takes place during modulation is diminished; see (10) and (11). Depression of both noise and distortion, but with greater emphasis upon the reduction of the latter, can be effected by raising the modulation level by an amount somewhat less than the feedback factor. This procedure has already been discussed in connection with (22) which gives the resulting noise-to-signal power ratio. Under similar conditions we have, from (25),

$$\sigma_F'' = \alpha AB\left(\frac{b_1}{F_2}\rho_1 S + \frac{1}{F_1}\left[\frac{b_2}{F_2^3}(\rho_1 S)^2 + \frac{1}{F_2^4}\left(b_3 - \frac{2b_2^2}{b_1}\right)(\rho_1 S)^3\right]\right) \quad (29)$$

when the feedback factor is large.

Equations (22) and (29) are most readily interpreted by means of Fig. 3, which illustrates the manner in which the receiver output is modified as the feedback is increased.

It is assumed that a constant signal level is maintained by an increase in modulation level up to a point corresponding to the factor F_1. Beyond this point the modulation remains fixed while the feedback

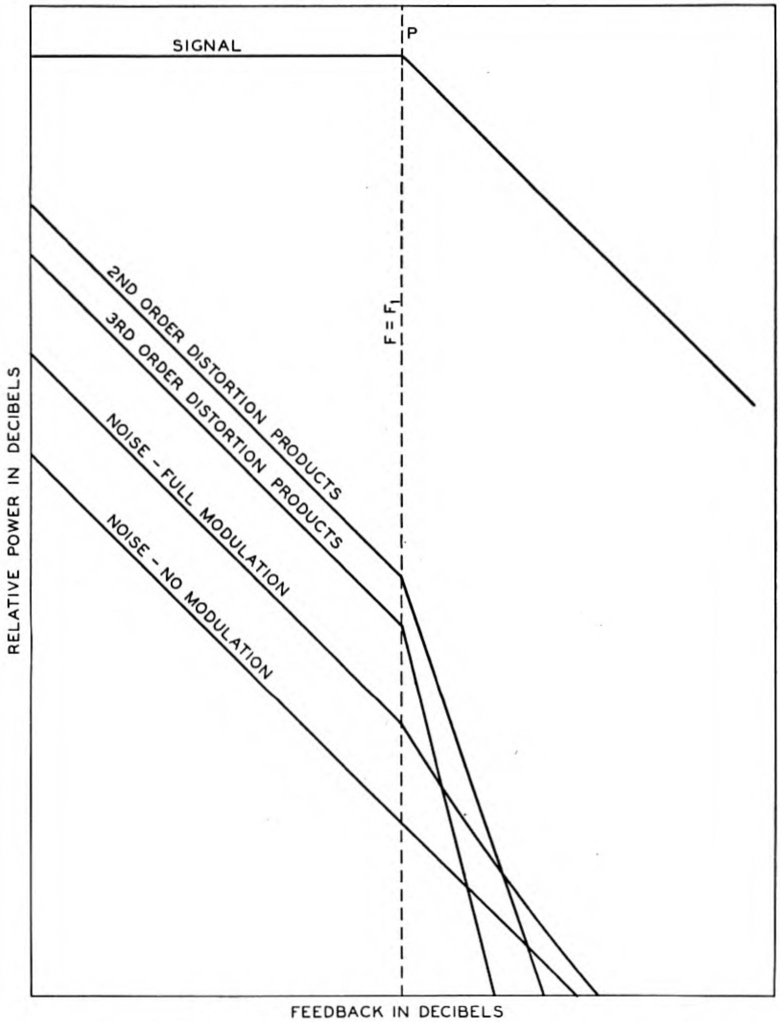


Fig. 3—Theoretical manner in which components of receiver output are modified by feedback. Modulation level at transmitter is assumed to be increased by the feedback factor up to point P, and subsequently to remain fixed.

factor is increased to F_1F_2. Since signal and noise levels have been expressed in terms of power, distortion levels are similarly expressed.

These levels, to an arbitrary decibel scale, have been plotted against decibels of feedback, given by the expression

$$10 \log F^2 = 20 \log F.$$

Balanced detection and the fulfillment of condition (28) have been assumed.

Over the region in which a constant signal output is maintained by increasing the modulation level, noise and distortion levels decrease in accordance with the feedback. The noise level during modulation continues to exceed the background noise by 4 decibels, assuming an initial frequency shift equal to the highest signal frequency to which the system is responsive.

Beyond the point at which the feedback factor has reached the value F_1, the modulation level at the transmitter is held constant. A further increase in feedback brings about a corresponding decrease in the effective percentage of modulation for the system, causing the signal level to fall in similar fashion. Distortion products now fall off still more rapidly with respect to the signal, so that an increase in feedback amounting to 1 decibel improves the second and third order distortion ratios by 2 and 3 decibels, respectively.

The ratio of signal to background or non-signaling noise remains fixed in this region in spite of the reduction in effective modulation. This ratio is that which would be obtained in a limiter system in which the same high-frequency band is transmitted. The noise increment, however, is diminished by the additional feedback and is made to approach zero.

By suitable choice of the variables F_1 and F_2 it is possible to proportion the benefits of feedback in the most advantageous manner. Thus if noise is of more consequence than distortion, modulation would be increased to the full extent of the feedback; if distortion is of primary concern, as it might well be in a multiplex system, operation as indicated in Fig. 3 would be preferable.

Experimental Results

Description of Equipment

Experimental confirmation of the principles which have been outlined has been obtained with the aid of a laboratory system shown schematically in Fig. 4. This arrangement provided a transmitter, receiver, and source of disturbance all under local control. The transmitter operated at a carried frequency of 20 megacycles. This was frequency-modulated by means of a circuit basically similar to that described by Travis.[7] Tube noise voltage appearing at the output of a high gain radio frequency amplifier supplied the high-frequency disturbance.

At the receiver an oscillator similar to that at the transmitter served to beat down the incoming wave to an intermediate frequency of 438 kilocycles. This was applied to a three-stage amplifier having substantially uniform gain over a band of 50 kilocycles, and thence delivered to a balanced frequency detector. In addition to signal voltage, automatic-frequency-control potentials were derived from the detectors. Both were carried back to the local oscillator, but in order to permit independent control of the amount of feedback their respective paths were kept separate. In this way full frequency control could be had even with signal feedback reduced to zero.

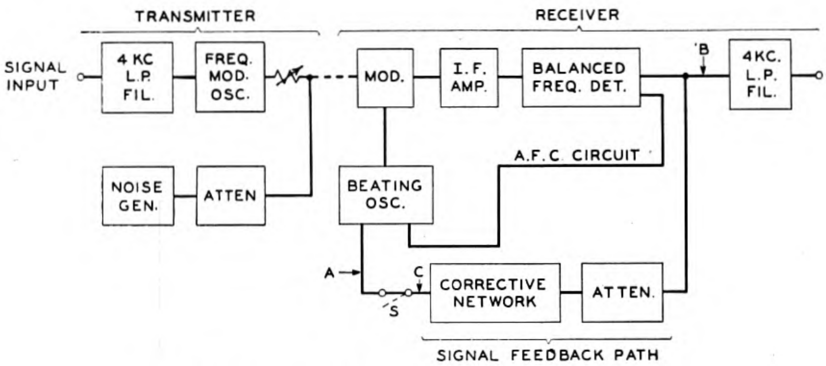


Fig. 4—Schematic of experimental feedback system.

Details of the frequency detector and feedback connections are shown in Fig. 5. The conversion system derives its characteristics from anti-resonant circuits L_1C_1 and L_2C_2, double-winding coils being used to isolate the rectifier anodes from the plate battery. One circuit is tuned to a frequency 15.4 kilocycles above the intermediate carrier frequency and the other to a corresponding point below, their characteristics intersecting at a point where the gain is approximately one half of the peak value. Detection takes place in linear rectifiers D_1 and D_2. By means of the arrangement shown, signal potentials are impressed upon the grids of amplifiers A_1 and A_2, while frequency-control voltage appears across condensers C_3, C_4. This voltage becomes zero when the receiver is correctly tuned and appears with proper polarity to

[7] Charles Travis, "Automatic Frequency Control," *Proc. I. R. E.*, vol. 23, pp. 1125–1141, October 1935.

produce correction of the frequency of the local oscillator in case of slow drifts in the frequency of either oscillator.

The use of a conversion system having peaks separated by an amount considerably exceeding the greatest frequency deviation is the result of a compromise between the readily adjustable and high impedance anti-resonant type of load circuit and others which, though more linear in their characteristics, lead to much lower gain in the conversion stage. While a peak separation of 14 kilocycles would have sufficed in view of the limitations of the transmitter, a considerably greater peak separation without corresponding increase in modulation was used. As a result that portion of the circuit characteristic actually embraced by the modulated intermediate-frequency wave

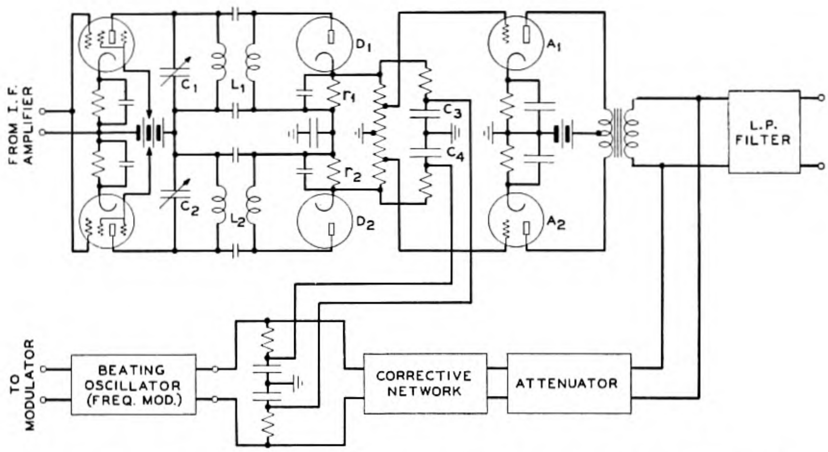


Fig. 5—Details of balanced frequency detector and feedback connections.

presented a much better approximation of a straight line than would have been possible with minimum peak separation. The penalty for adjusting the circuits in this manner is merely a loss in detecting efficiency and not an impairment of the signal-to-noise ratio. This can readily be overcome by additional audio-frequency amplification.

The signal-frequency feedback path includes an attenuator for adjusting the feedback and a corrective network for preventing singing around the feedback loop. Frequency control and feedback paths are finally combined at the modulation terminals of the local oscillator.

The necessity for the inclusion of a corrective network to modify the transmission characteristics of the feedback path is evident from Fig. 6. This shows the measured gain and phase characteristics of the receiver alone, viewed as a voice-frequency network between points A and B in Fig. 4. This was obtained by applying signal frequencies to the modulation terminals of the beating oscillator and making observations at point B with switch S open, proper termination being provided at both sides of the break. The unmodulated transmitter served, in effect, as the beating oscillator during this measurement. At the lower signal frequencies the phase is practically 180 degrees as indicated by (4) with $x_1 = 0$. As the signal frequency is increased the phase

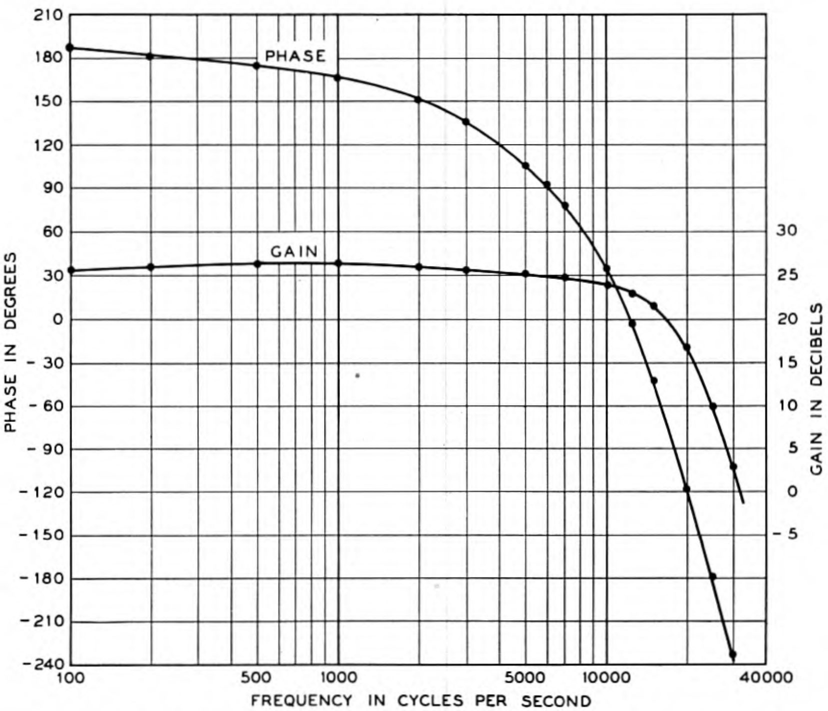


Fig. 6—Phase and gain characteristics of receiver measured between points A and B of Fig. 4 with switch S open. Transmitter in operation but not modulated.

is progressively shifted from this value. Except for that produced by the output transformer, the shift takes place within the intermediate-frequency amplifier and conversion circuits. Its magnitude is a measure of the slope of the phase-frequency characteristic of the intermediate frequency system.

The existence of positive gain at a point of zero phase shows that singing would occur if feedback connections were made directly to the beating oscillator. It was therefore necessary to reduce the gain below unity at the point of zero phase. This was accomplished by including in the feedback path a network designed by R. L. Dietzold. The gain-frequency characteristic of this network is shown in Fig. 7. The modified loop characteristics as measured between points A and C with switch S open, and with the attenuator set for an 8-decibel loss, are given in Fig. 8. Full feedback is applied only over a band extending to 4 kilocycles, so that the range of frequencies applied to the transmitter and delivered to the listener must be restricted to this figure.

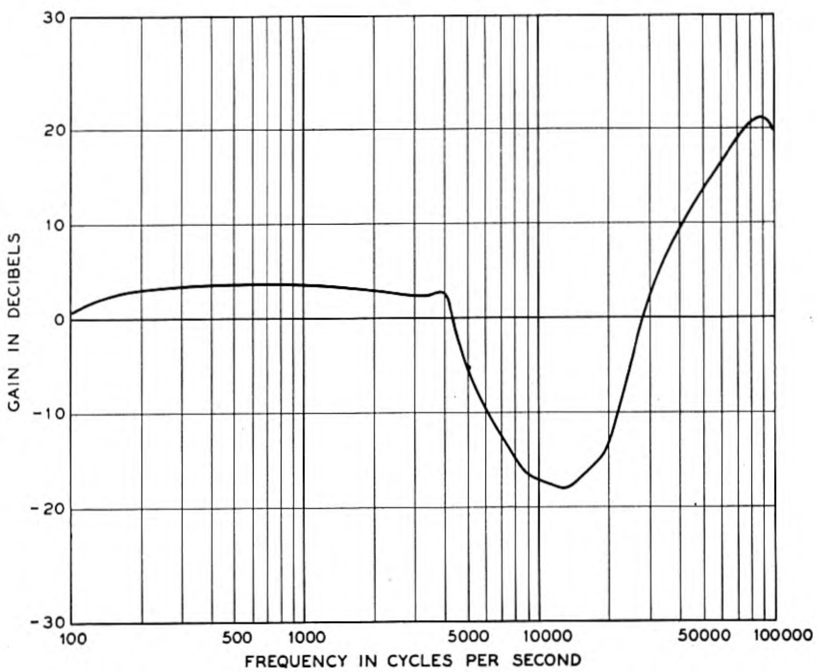


Fig. 7—Gain-frequency characteristic of corrective network inserted in signal feedback path.

The limit of stable feedback which can be realized is indicated by the difference between the loop gain within the useful band and that at the frequency corresponding to zero phase.

Distortion Measurements

The manner in which distortion levels at the output of the receiver were observed to vary with feedback is depicted in Figs. 9 to 12. In each case the modulation level for zero feedback was such as to shift the frequency of the transmitter ± 7 kilocycles at the rate of 1000 cycles per second. Figure 9 shows the effect of increasing the modulation in proportion to the feedback so as to maintain a constant output level for the fundamental. Both second and third harmonics tend to be reduced in proportion to the feedback, the improvement in third-harmonic level being 23.5 decibels for 25-decibel feedback. Failure to realize full reduction of the second harmonic is the result of distortion beginning to manifest itself in one or the other of the modulated oscillators. At the point of 25-decibel feedback the transmitter and

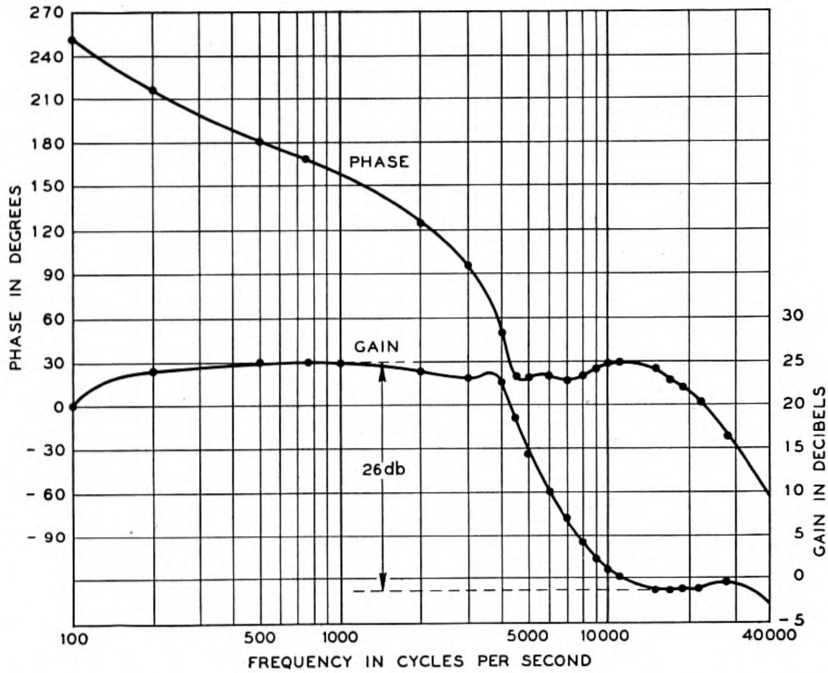


Fig. 8—Gain and phase characteristics of complete feedback loop including corrective network. Measured between points A and C of Fig. 4 with switch S open.

beating oscillator were being modulated to the extent of ± 124.5 and ± 117.5 kilocycles, respectively.

The curves of Fig. 10 were obtained by maintaining a constant fundamental level up to the point of 15-decibel feedback and then allowing the modulation level at the transmitter to remain constant thereafter. The results correspond rather closely with the theoretical curves of Fig. 3 and show the very rapid decrease in distortion which takes place when the modulation level remains unaltered; see (25). A more extreme example of this method of operation is shown in Fig. 11 where modulation was left at its initial value. Harmonic levels soon reached a point beyond which they could not be measured accurately.

In a practical system the loss in signal level resulting from operation in this manner could easily be overcome by the addition of a low-distortion audio-frequency amplifier at the output of the receiver. This amplifier might well embody negative feedback of the more usual type.

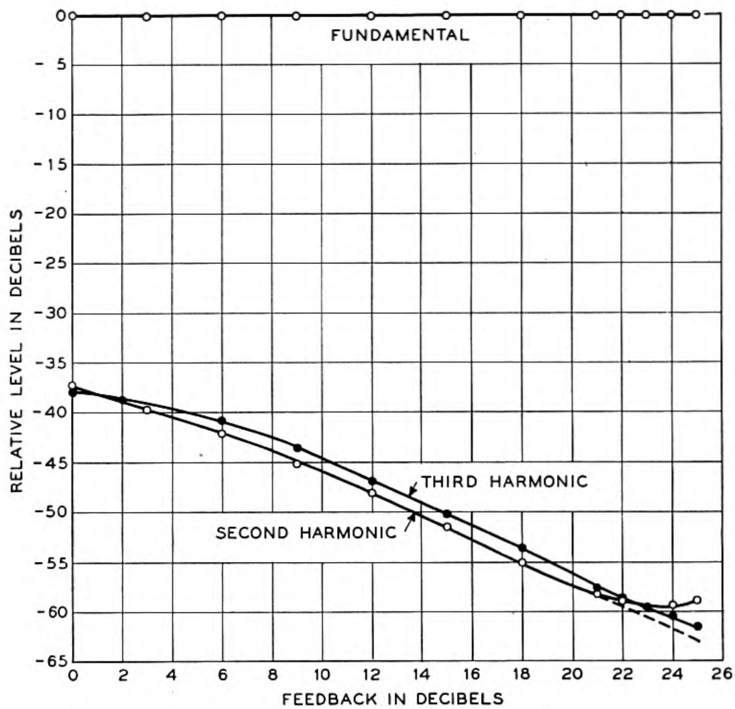


Fig. 9—Effect of feedback upon receiver distortion. Fundamental level kept constant by increasing transmitter modulation in proportion to the feedback. Modulation with no feedback = ± 7 kilocycles at 1000 cycles per second.

A composite of these distortion measurements is given in Fig. 12. Harmonic levels are plotted in decibels below the fundamental and are indicative of the improvements brought about by feedback. If it is assumed that any loss in signal is compensated by additional audio-frequency amplification, the fundamental level would be represented in all cases by the axis of abscissae.

Noise Measurements

In Fig. 13 are given the results of a series of observations of receiver output noise versus amount of feedback for a number of high-frequency disturbance levels. Measurements were made in the absence of modulation and hence are indicative of the manner in which background noise is modified by feedback. The signal level indicated is that which could be maintained at low noise levels by increasing the modulation in proportion to the feedback, and is not significant for observations falling within or close to the shaded area, as will be explained subsequently.

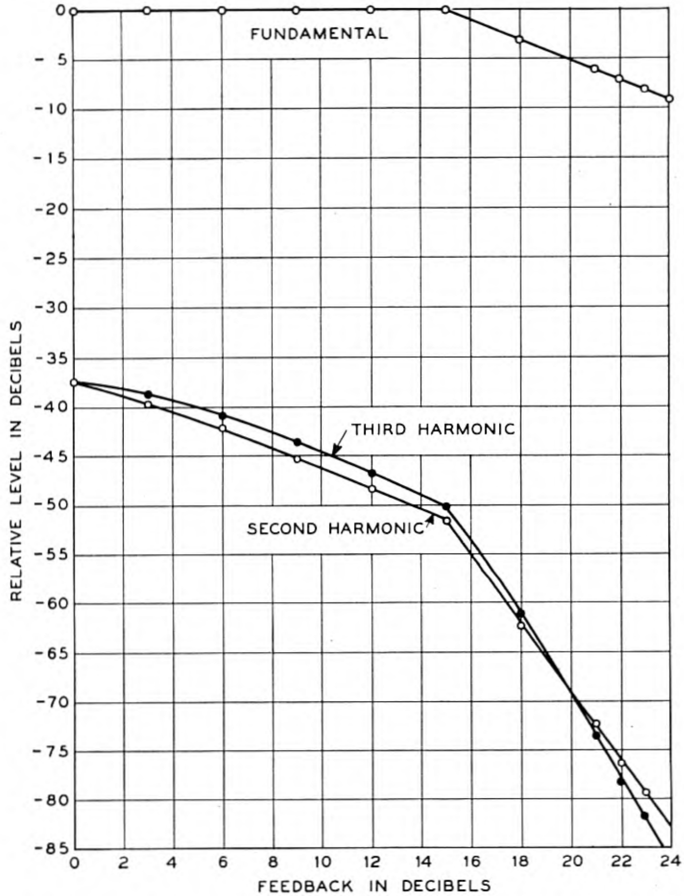


Fig. 10—Effect of feedback upon receiver distortion. Conditions same as indicated for Fig. 9 up to 15-decibel feedback; modulation held constant thereafter.

The lowest noise level shown is that generated within the receiver while the higher levels were produced by disturbances introduced from the noise generator. The relative magnitude of the effective carrier and disturbing voltages at the grids of the amplitude detector is indicated on each curve. This was obtained in the following manner: With the transmitter turned off the noise attenuator was adjusted until the introduced disturbance produced the same value of rectified current as that observed when the carrier alone was applied. This determined the input level from the noise generator which produced equal root-

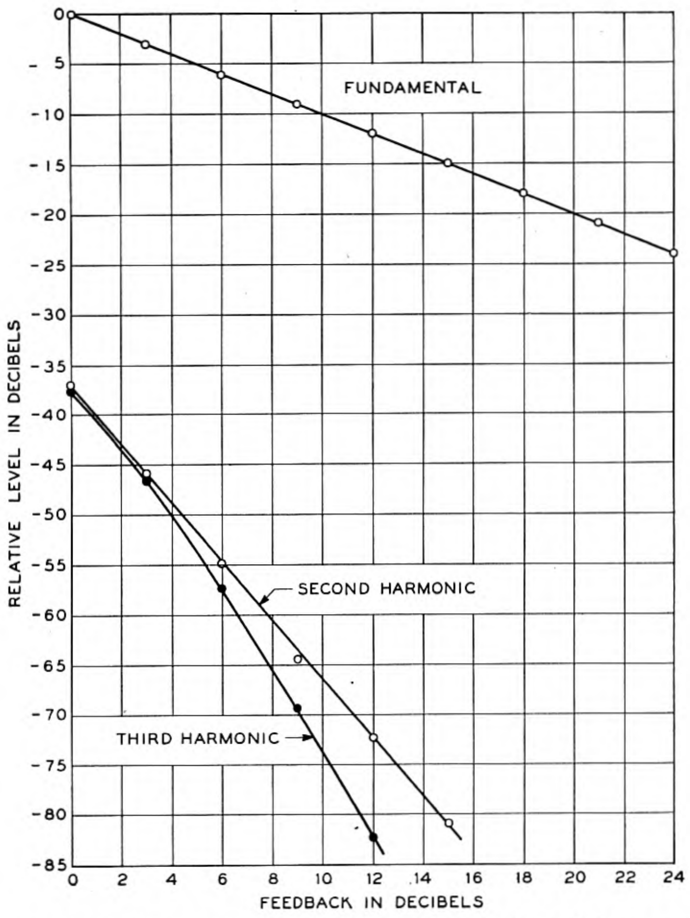


Fig. 11—Effect of feedback upon receiver distortion. Modulation held to a constant value of $\pm$ 7 kilocycles.

mean-square values of intermediate-frequency carrier and disturbance. Since at very low inputs from the noise generator the net intermediate-frequency disturbance was determined partly by tube noise generated within the receiver, a curve of output noise without feedback versus input from the noise generator was obtained. In the region where the effect of receiver tube noise was evident the assumption of a linear relationship between disturbance level and output noise permitted correction of the curve so that equivalent disturbance levels could be related to any setting of the noise attenuator, or to the receiver output noise level without feedback.

The signal-to-noise ratios at the output of the receiver without feedback are considerably higher than the corresponding ratios of carrier and disturbance levels existing at the amplitude detectors. This is the

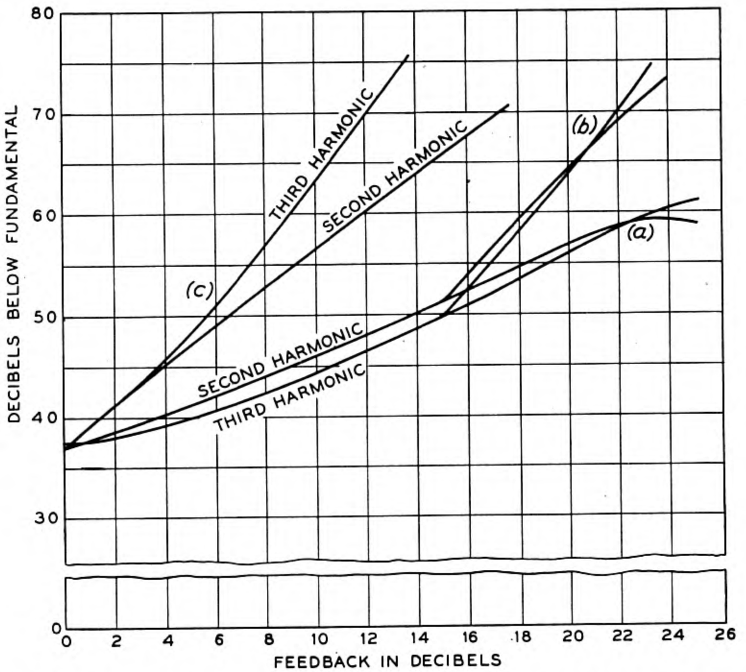


Fig. 12—Composite of data from Figs. 9 to 11 expressing ratios of harmonic levels to fundamental level. Curves (*a*) from Fig. 9, (*b*) from Fig. 10, and (*c*) from Fig. 11.

result of two factors. The intermediate-frequency wave is modulated to the extent of $\pm$ 7 kilocycles while the range of frequencies appearing at the output terminals of the receiver is limited to 4 kilocycles. Hence at the output of the balanced detector the noise level in the absence of modulation is 9.6 decibels below that which would be observed at the output of an amplitude-modulation system. Furthermore the admittance characteristic of the complete intermediate-frequency system is such that the effective disturbing voltage delivered to the amplitude detectors is 11 decibels greater than that admitted by an amplitude modulation system having the minimum intermediate-frequency band width of 8 kilocycles.

Aural observation of the character of the output noise showed that, excluding the shaded area in Fig. 13, the normal characteristics of fluctuation noise are preserved as feedback is applied. Upon crossing

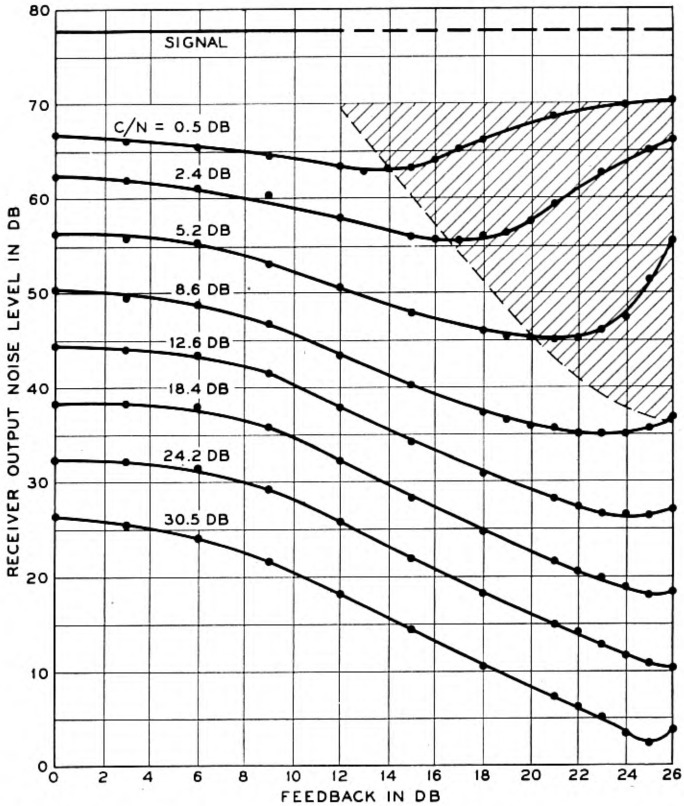


Fig. 13—Effect of feedback upon receiver output noise level for various amounts of high-frequency disturbance. Ratio of root-mean-square values of carrier and disturbance is shown on each curve. Shaded area indicates region of "crackling" in the absence of modulation.

the boundary of the shaded area the noise becomes punctuated with intermittent clicks which increase in rapidity and violence as the feedback factor is raised, giving rise to what can be described as "crackling." After passing through a region of maximum turbulence the noise gradually assumes the nature of a much higher level of fluctuation noise.

The region embracing the appearance of the above phenomenon is also characterized by a marked reduction in signal level. At high modulation levels "cracking" begins at a somewhat lower disturbance level than is necessary to initiate it in the absence of modulation. The initial effect is to impart a roughness to tone modulation. Further increase in disturbance level produces a rapid depression of the signal, so that it soon becomes submerged in noise. The manner in which this depression takes place is shown in Fig. 14. The signal, produced by 1000-cycle modulation was measured by means of a highly selective

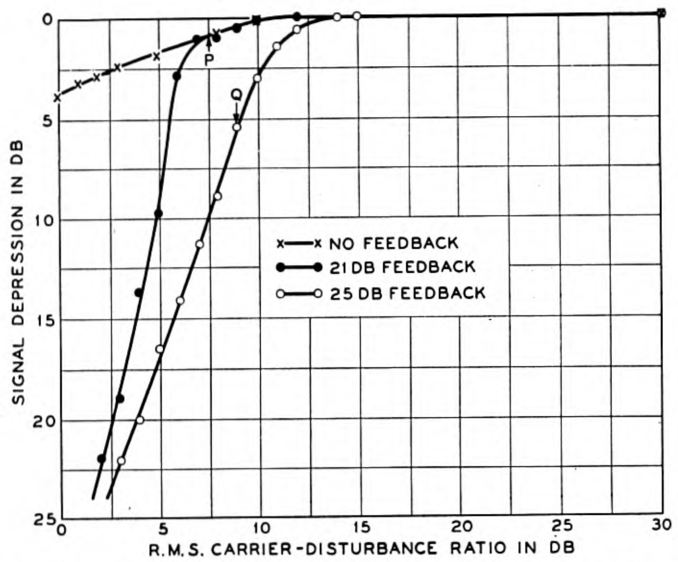


Fig. 14—Depression of output signal by the disturbance. Modulation: ± 7 kilocycles for no feedback, ± 78.5 kilocycles for 21-decibel feedback, and ± 124.5 kilocycles for 25-decibel feedback. Points P and Q denote incidence of crackling in the absence of modulation, for 21- and 25-decibel feedback, respectively.

analyzer so that observations could be carried well below the general noise level.

The point at which depression of the signal begins coincides with the appearance of roughness in the output tone resulting from the momentary suppression of the signal by the higher noise peaks. A further increase in disturbance level increases the number of peaks per second which rise above the critical value, and the energy content of the signal is rapidly diminished. The point at which faint crackling could first be detected in the absence of modulation is indicated on each curve.

The signal-to-noise ratios obtained with zero and with 25 decibels of feedback are shown in Fig. 15. These are plotted against the ratio of root-mean-square carrier and disturbance levels at the end of the intermediate-frequency channel. Signal levels were measured in the presence of the disturbance so as to take into account the depression of the signal, while noise levels were observed in the absence of modulation.

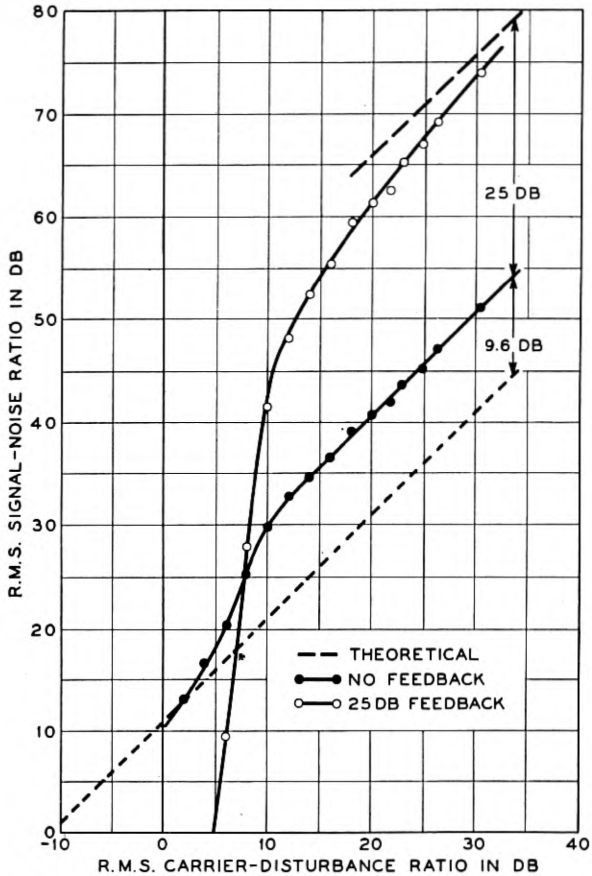


Fig. 15—Output signal-to-noise ratio vs. carrier-disturbance ratio with and without feedback. Modulation ± 124.5 kilocycles for 25-decibel feedback and ± 7 kilocycles for no feedback.

The improvement resulting from the application of feedback is given by the difference between the two curves and approaches the theoretical improvement at the low noise levels. The curve obtained with feedback exhibits a rather sharp break when the ratio of carrier to effective disturbance is in the vicinity of 10 decibels. Experimental data [6] published by Crosby indicates that in the case of fluctuation noise the ratio of the maximum peak amplitude to the root-mean-square value is about 13 decibels. The corresponding figure in the case of a sine wave is 3 decibels. Hence equality of carrier peak amplitude and the highest peaks of the disturbance obtains when the ratio of their root-mean-square values is 10 decibels. With feedback this condition appears to define a fairly critical disturbance level above which the output signal-to-noise ratio is very rapidly diminished. Crosby has shown [6] that with systems employing amplitude limiters a similar condition marks the point beyond which the noise improvement realized at the lower disturbance levels is soon lost. This point he has termed the "threshold of noise improvement."

A less sharply defined break also occurs in the curve expressing noise conditions in the absence of feedback. This is the result of a progressive destruction, at the higher disturbance levels, of the balancing out of amplitude effects in the push-pull detector which is realized when the noise is low.

While direct comparison of the feedback system with an actual amplitude modulation system was not possible with the equipment used, it is thought that a comparison based upon theoretical considerations may be of interest. The procedure is as follows: The noise ratios shown in Fig. 15 for the system without feedback are, for disturbances below the threshold value, 9.6 decibels in excess of those which would be realized in a fully modulated amplitude system. A dotted line, displaced from the linear portion of the measured curve by this amount, is shown. The abscissae of the dotted curve do not represent the true carrier-disturbance ratio which would obtain in the amplitude system for the reason that, ideally, the intermediate-frequency amplifying system would have a band width of but 8 kilocycles. In such a system the signal-to-noise ratio would be equal to the carrier-disturbance ratio except at the very high noise levels. Hence the intercept of the dotted line with the axis of abscissas marks the point of equal carrier and disturbance levels in this system. The difference of 11 decibels between this point and the zero point on the scale as drawn measures the amount by which the disturbance at the rectifiers in the experimental system exceeds that which would be found in the ideal amplitude system.[8] Consequently, if it is desired to relate the data of Fig. 15 to the disturbance ratio which would exist at the input to the detector in the amplitude system, and hence to the signal-to-noise ratio in that

[8] Comparison of the areas under idealized curves representing the square of the transmission through the intermediate-frequency systems in the two cases indicates a difference of 10.1 decibels.

system, it is merely necessary to displace the experimental curves to the right by 11 decibels.

Figure 16 shows such a comparison between the feedback system adjusted to give 25 decibels of feedback, and an ideal amplitude modula-

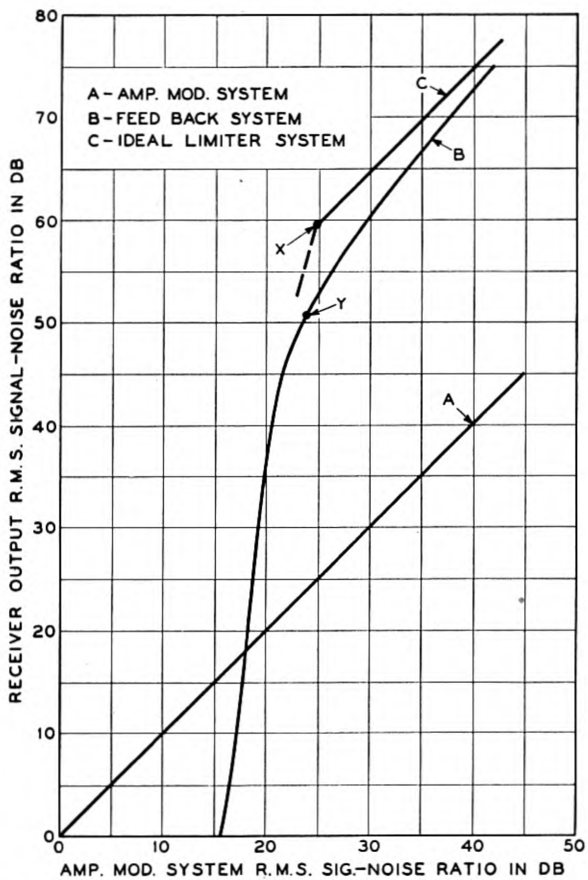


Fig. 16—Theoretical comparison of signal-to-noise ratios obtained at 25-decibel feedback (curve B) with amplitude-modulated system (curve A) and ideal limiter system with deviation ratio = 31.1 (curve C). Point X = threshold of noise improvement for limiter system. Point Y = point where "crackling" first became evident in the presence of ± 124.5-kilocycle modulation.

tion system. There is also included a curve showing the theoretical performance which would be approached by a frequency-modulation system using amplitude limitation. Transmitted band width and audio-frequency response equal to that used in the experimental feedback system have been assumed. This corresponds to a deviation ratio of 124.5 kilocycles ÷ 4 kilocycles = 31.1, resulting in a theoretical noise deduction of 34.6 decibels at low disturbance levels. The threshold of noise improvement, indicated at the point x, is located at a point where the peak signal-to-noise ratio in the amplitude system is equal to the square root of the deviation ratio.[6] This factor takes account of the higher disturbance level in the intermediate-frequency channel of the wide-band system. Assuming a factor of 10 decibels between maximum peak and root-mean-square values of fluctuation noise this corresponds to a root-mean-square signal-to-noise ratio of 24.9 decibels in the amplitude system.[9] In the feedback system the point at which crackling was first observed in the presence of modulation is shown at Y. This coincides very closely with the theoretical threshold of noise improvement in the limiter system.

Conclusions

It has been shown that the application of negative feedback to a frequency-modulation system affords a means for effecting large reductions in both noise and receiver distortion. The theoretical analyses of these effects, while they have been simplified to an extent which makes them inadequate to cover all conditions which can be encountered in practice, are adequately substantiated by the observed performance of the experimental system within the limitations of the theory.

Substantial benefits are realized only when the amount of feedback is large, and when the disturbance level is not too great. In common with frequency-modulation systems employing amplitude limitation a large reduction in noise must be paid for by increasing the band width of the transmitted wave. While the principles involved in the two systems are quite different, their performance as regards noise modification, both at high and at low levels of disturbance, exhibits striking similarities. The ability to reduce distortion is, on the other hand, a feature found only in the feedback system.

Acknowledgment

The writer wishes to acknowledge his indebtedness to the following of his colleagues: To Dr. H. W. Bode for his investigations of the problem of stability in feedback systems, and to Mr. R. L. Dietzold for the design of the stabilizing network which was used; to Mr. W. R. Bennett whose unpublished work on basic frequency-modulation prob-

[9] In the absence of more exact information regarding the performance of a system using this large a deviation ratio the portion of this curve below the threshold has been omitted.

lems has been of great value; and to Messrs. E. A. Krauth and O. E. DeLange for assistance in the experimental work. Reference has already been made to the theoretical work of Dr. J. R. Carson.

APPENDIX A

Analysis of Distortion Reduction

Assume that the transmitter is frequency-modulated with a signal wave which we shall represent by the symbol $S = S(t)$. Then the instantaneous frequency of the transmitter will be

$$\omega_1 + \rho_1 S. \tag{30}$$

The instantaneous phase of the transmitted wave is the integral of this expression and the voltage delivered to the input of receiver can be written

$$A \cos\left(\omega_1 t + \rho_1 \int_0^t S dt + \phi_1 \right). \tag{31}$$

Designating the low-frequency voltage delivered at the output of the receiver as $\sigma = \sigma(t)$ the result of feeding back a portion of $k\sigma$ of the output so as to frequency-modulate the local oscillator is the wave

$$B \cos\left(\omega_2 t + \rho_2 \int_0^t k\sigma dt + \phi_2 \right). \tag{32}$$

Application of these two waves to the modulator produces the intermediate-frequency product [10]

$$\alpha AB \cos\left[\omega_0 t + \rho_1 \int_0^t S dt - \rho_2 \int_0^t k\sigma dt + \phi_0 \right] \tag{33}$$

where

$$\omega_0 = \omega_1 - \omega_2$$
$$\phi_0 = \phi_1 - \phi_2.$$

Terms in the above which involve the integral sign represent phase angles which vary with time. Hence we shall rewrite (33) more compactly

$$\alpha AB \cos\left[\omega_0 t + \theta(t) + \phi_0\right]. \tag{34}$$

It has been shown by Carson and Fry [3] that the process of detecting a frequency-modulated wave is, in effect, its differentiation. Since the high-frequency wave itself exhibits the integral of the signal wave, see

[10] This expression constitutes a more general form of (4).

(30), it can be reasoned that a differentiation process is necessary for the recovery of the signal itself.

Differentiation of the argument of the cosine term in (31) yields the instantaneous frequency (rate of change of phase with respect to time) of the received wave given by (30). Now it can be shown that with a strictly linear frequency detector, the low-frequency output is proportional to the response of the conversion system at the instantaneous frequency. The recovered signal is, therefore, proportional to the variable part of the instantaneous frequency and hence to the time derivative of the variable phase term in the original wave.

In the case of non-linearity in the characteristic of the frequency detector the output can be expressed, to a sufficiently close degree of approximation, as a power series in the derivative of the phase term $\theta(t)$. Hence the output of the receiver can be written in the form

$$\sigma(t) = \alpha AB \sum_n b_n \left[\frac{d}{dt} \theta(t) \right]^n \tag{35}$$

$$= \alpha AB \sum_n b_n [\rho_1 S - k\rho_2 \sigma]^n \tag{36}$$

where the coefficients b_n are based upon the transfer admittance characteristic of the receiver.

What is now desired is the relationship between σ and S. This can be expressed in the general form

$$\sigma = \alpha AB \sum_n c_n [\rho_1 S]^n. \tag{37}$$

Equations (36) and (37) can now be equated. Replacing σ in the right-hand side of (36) by the series (37) we shall have

$$\begin{aligned} c_1\rho_1 S &+ c_2(\rho_1 S)^2 + c_3(\rho_1 S)^3 + \cdots \\ &= b_1[\rho_1 S - k\alpha AB\rho_2(c_1\rho_1 S + c_2(\rho_1 S)^2 + c_3(\rho_1 S)^3 + \cdots)] \\ &\quad + b_2[\rho_1 S - k\alpha AB\rho_2(c_1\rho_1 S + c_2(\rho_1 S)^2 + c_3(\rho_1 S)^3 + \cdots)]^2 \\ &\quad + b_3[\rho_1 S - k\alpha AB\rho_2(c_1\rho_1 S + c_2(\rho_1 S)^2 + c_3(\rho_1 S)^3 + \cdots)]^3 \\ &\quad + \cdots . \end{aligned} \tag{38}$$

After expanding, coefficients of like powers of $\rho_1 S$ can be equated. Then solving for the first three orders of c_n we find

$$c_1 = \frac{b_1}{1 + \alpha b_1 kAB\rho_2} = \frac{b_1}{1 - \mu\beta} \tag{39}$$

$$c_2 = \frac{b_2}{(1 - \mu\beta)^3} \tag{40}$$

$$c_3 = \frac{b_3}{(1 - \mu\beta)^4} - \frac{2\alpha ABk\rho_2 b_2{}^2}{(1 - \mu\beta)^5} \tag{41}$$

where

$$\mu = \alpha ABb_1 \qquad \text{and} \qquad \beta = -k\rho_2.$$

Inserting these values in (37) and writing $(1 - \mu\beta) = F$, the receiver output becomes, with feedback,

$$\sigma_F = \alpha AB\left[\frac{b_1}{F}\rho_1 S + \frac{b_2}{F^3}(\rho_1 S)^2 + \left(\frac{b_3}{F^4} - \frac{2b_2^2}{b_1}\cdot\frac{F-1}{F^5}\right)(\rho_1 S)^3\right]. \tag{42}$$

When $F \gg 1$ this can be written

$$\sigma_F = \alpha AB\left[\frac{b_1}{F}\rho_1 S + \frac{b_2}{F^3}(\rho_1 S)^2 + \frac{1}{F^4}\left(b_3 - \frac{2b_2^2}{b_1}\right)(\rho_1 S)^3\right]. \tag{43}$$

Without feedback we have

$$\sigma = \alpha AB[b_1\rho_1 S + b_2(\rho_1 S)^2 + b_3(\rho_1 S)^3]. \tag{44}$$

APPENDIX B

Analysis of Noise Reduction

In the following analysis it is assumed that the amplitude of the disturbance producing the noise is sufficiently small compared with that of the incoming signal wave so that the principle of superposition will apply. Hence the manner in which the effect of a single disturbing component is modified by feedback will first be developed. Then the effect of a disturbance consisting of a continuous spectrum is derived by direct summation.

Consider first the case where there are impressed upon the grid of the modulator the incoming wave and the local oscillator voltage [11] as defined by (31) and (32), plus a single disturbing component

$$Q \cos\left[(\omega_1 + \omega_n)t + \phi_n\right]. \tag{45}$$

Then the intermediate-frequency product will be

$$\alpha AB \cos\left[\omega_0 t + \rho_1\int_0^t S\,dt - \rho_2\int_0^t k\sigma\,dt\right] + \alpha BQ\cos\left[(\omega_0 + \omega_n)t - \rho_2\int_0^t k\sigma\,dt + \phi_n\right]. \tag{46}$$

For simplicity assume that the intermediate-frequency amplifier and conversion circuit have the ideal transfer admittance characteristic

$$Y(\omega) = a_0 + a_1(\omega - \omega_0). \tag{47}$$

[11] Arbitrary phase constants will be omitted from these expressions since they do not affect the final result.

Then all derivatives of Y with respect to ω above the first are zero, and the steady-state response is equal to its response at the instantaneous frequency of the applied wave. Hence after conversion we shall have

$$\alpha AB[a_0 + a_1(\rho_1 S - \rho_2 k\sigma)] \cos\left[\omega_0 t + \int_0^t (\rho_1 S - \rho_2 k\sigma)dt\right]$$
$$+ \alpha BQ[a_0 + a_1(\omega_n - \rho_2 k\sigma)] \cos\left[(\omega_0 + \omega_n)t - \int_0^t \rho_2 k\sigma dt + \phi_n\right]. \quad (48)$$

Application of (48) to a linear amplitude detector will yield a low-frequency output proportional to its amplitude. The amplitude factor is readily calculated for the case where $AB \gg BQ$. For if

$$X \cos x + Y \cos y = Z \cos z$$

then

$$Z = \sqrt{X^2 + Y^2 + 2XY \cos (x - y)}$$

and when $X \gg Y$

$$Z \doteq X + Y \cos (x - y). \quad (49)$$

Hence the output of the linear detector will be

$$\gamma\left(\alpha AB[a_0 + a_1(\rho_1 S - \rho_2 k\sigma)] + \alpha BQ[a_0 + a_1(\omega_n - \rho_2 k\sigma)]\right.$$
$$\left.\times \cos\left[\omega_n t - \rho_1 \int_0^t S dt + \phi_n\right]\right). \quad (50)$$

The term $\alpha\gamma ABa_0$ represents direct current. Assuming that this is not fed back to the local oscillator we can then write

$$\sigma = A'[a_1(\rho_1 S - \rho_2 k\sigma)] + Q'[a_0 + a_1(\omega_n - \rho_2 k\sigma)] \cos \xi \quad (51)$$

where

$$A' = \alpha\gamma AB$$
$$Q' = \alpha\gamma BQ$$
$$\xi = \left(\omega_n t - \int_0^t \rho_1 S dt + \phi_n\right).$$

Solving for σ

$$\sigma = \frac{1}{1 + a_1 A' k\rho_2}\left[1 + \frac{a_1 Q' k\rho_2 \cos \xi}{1 + a_1 A' k\rho_2}\right]^{-1}$$
$$\times [A' a_1 \rho_1 S + Q'(a_0 + a_1 \omega_n) \cos \xi]. \quad (52)$$

If $Q' \ll A'$

$$\sigma \doteq \frac{1}{F}\left[1 - \frac{a_1 Q' k\rho_2 \cos \xi}{F}\right][A' a_1 \rho_1 S + Q'(a_0 + a_1\omega_n) \cos \xi] \quad (53)$$

where $F = 1 + a_1 A' k\rho_2 = 1 - \mu\beta$ as before. Finally, neglecting terms in Q'^2, we have

$$\sigma = \frac{1}{F}\left[A' a_1 \rho_1 S + Q'\left(a_0 + a_1\omega_n - \frac{A' a_1^2 \rho_1 k \rho_2 S}{F}\right)\cos \xi\right]. \quad (54)$$

The first term is the recovered signal while the remaining terms represent noise. Both signal and noise are modified by feedback. If we let

$$\rho_1 S = \Delta\omega \cos pt \quad (55)$$

then the noise becomes

$$\frac{Q'}{F}\left[(a_0 + a_1\omega_n) - \frac{1}{F}(A' a_1^2 k\rho_2 \Delta\omega \cos pt)\right] \times \cos(\omega_n t - x \sin pt + \phi_n). \quad (56)$$

By means of the Jacobi expansions this can be put in the form

$$\frac{Q'}{F}\sum_{m=-\infty}^{\infty}\left[(a_0 + a_1\omega_n) - \frac{A' a_1^2 k\rho_2 mp}{F}\right] J_m(x) \times \cos[(\omega_n - mp)t + \phi_n] \quad (57)$$

where $J_m(x)$ is the Bessel coefficient of the first kind.

Now let it be assumed that the disturbance consists of a very large number of sinusoidal components of like amplitude Q, random phase, and uniformly distributed along the frequency scale. The summation of this series of voltages can be represented by the very general expression

$$f(t) \cos[\omega t + \phi(t)]. \quad (58)$$

So long as $f(t)$, the equivalent amplitude of the high-frequency disturbance, is small compared with the carrier amplitude A, the approximation (49) will be valid and the total output noise can be obtained by summing up the effects of the individual elements which constitute the disturbance.

The effect of a single disturbing element is given by (57). Any term of this expression can be made to have the frequency q if m and ω_n are so chosen that

$$\omega_n = mp \pm q. \quad (59)$$

Then for each value of m in (57) there will be available values of ω_n

to satisfy both of the conditions expressed by (59). The total effect is obtained by summing the output power resulting from each contribution since the original elements have random phases. If r_2 is the resistance of the output circuit the total power of frequency q becomes

$$\frac{Q'^2}{2r_2F^2}\left(\sum_{m=-\infty}^{\infty}\left[a_0+a_1(mp+q)-\frac{A'a_1^2k\rho_2mp}{F}\right]^2 J_m^2(x)\right.$$
$$\left.+\sum_{m=-\infty}^{\infty}\left[a_0+a_1(mp-q)-\frac{A'a_1^2k\rho_2mp}{F}\right]^2 J_m^2(x)\right). \quad (60)$$

This is readily evaluated with the aid of tables appended to an earlier paper.[12] The result is

$$\frac{Q'^2}{r_2F^2}\left[a_0^2+\frac{a_1^2\Delta\omega^2}{2F^2}+a_1^2q^2\right]. \quad (61)$$

The amplitude factor Q remains to be defined. If N^2 is the mean disturbing power per unit band width in the vicinity of the carrier frequency and r_1 the resistance of the input circuit, the peak amplitude of any element is defined by the relation

$$N^2d\omega=\frac{Q^2}{2r_1}\cdot \quad (62)$$

Thus the power associated with each element becomes differentially small, and if the value so obtained is entered into (61) there is obtained the output noise power contained in a band extending from q to $q+dq$. Then we shall have

$$dW=\frac{2N^2r_1(\alpha\gamma B)^2}{r_2F^2}\left[a_0^2+\frac{a_1^2\Delta\omega^2}{2F^2}+a_1^2q^2\right]dq. \quad (63)$$

The total noise power in a band extending to a limiting frequency q_a is

$$P_n=\int_0^{q_a}dW=\frac{2N^2r_1(\alpha\gamma B)^2}{r_2F^2}\left[a_0^2+\frac{a_1^2\Delta\omega^2}{2F_2}+\frac{a_1^2q_a^2}{3}\right]q_a. \quad (64)$$

The corresponding signal power is

$$P_s=\frac{(\alpha\gamma AB)^2}{2r_2F^2}a_1^2\Delta\omega^2. \quad (65)$$

[12] J. G. Chaffee, "The Detection of Frequency Modulated Waves," *Proc. I. R. E.*, vol. 23, pp. 517–540, May 1935.

Survey of Magnetic Materials and Applications in the Telephone System

By V. E. LEGG

The great diversity of magnetic characteristics demanded by telephone apparatus, and the large number of available magnetic materials propose intricate problems in the choice of materials and design of apparatus to attain greatest efficiency and economy. The present paper undertakes to evaluate magnetic materials in relation to apparatus needs. After a review of the earlier developments, the materials now available are listed, together with data on technical characteristics and raw materials costs. The advantages of various materials for specific applications are described. The scope of possible further improvements in magnetic materials is surveyed.

Historical

TWENTY years ago, the telephone system used primarily iron, together with a small amount of silicon iron, for applications requiring soft magnetic materials, and carbon, tungsten or chromium steel for permanent magnet applications. The permalloys [1] were already fairly thoroughly developed by 1920 in what is now the Bell Telephone Laboratories, and 78.5 permalloy [2] shortly attained commercial recognition for its utility as a continuous loading material for submarine telegraph cables.[3] This and other nickel-iron alloys were soon serving in many types of telephone relays, and in various coils where the designs could be profitably modified to adapt them to the new materials. Upon the development of commercial means for embrittling and pulverizing permalloy, this material was soon in extensive use because it offered improved characteristics over the compressed powdered iron core material previously in use. Redesigns of filter and loading coils have introduced such economies that practically all these coils made by the Western Electric Company have until recently employed compressed powdered permalloy cores.[4]

A desire to reduce the losses in a-c. apparatus arising from eddy currents in magnetic parts led to the development of permalloys of higher

[1] H. D. Arnold & G. W. Elmen, *Jour. Frank. Inst. 195*, 621 (1923).
[2] The approximate chemical compositions of the various materials herein discussed are given in Tables I and II.
[3] O. E. Buckley, *Jour. A. I. E. E. 44*, 821 (1925).
[4] W. J. Shackelton & I. G. Barber, *Trans. A. I. E. E. 47*, 429 (1928).

resistivity, containing several per cent chromium or molybdenum.[5] The problems of embrittlement and pulverization of molybdenum permalloy were also successfully solved. This material has been recently adopted for manufacture of filter and loading coil cores,[6] in which material of higher resistivity is especially advantageous.

Attempts to decrease the losses due to hysteresis led to the discovery of the nickel-iron-cobalt alloys—the perminvars. A molybdenum-perminvar was perfected for use in the continuous loading of submarine telephone cable.[7]

The large economic advantages promised by improvements in soft magnetic materials confined much of the earlier work to this field. However, within the last 20 years new permanent magnet materials have been discovered here and abroad which offered radical improvements in this direction. Such materials have been introduced in telephone apparatus wherever found advantageous.

Characteristics of Available Materials

The number of different magnetic materials in use is quite large on account of the various combinations of properties required for special applications and on account of a multiplicity of trade names. For the present purpose, an abbreviated listing is given of typical materials covering the whole range of magnetic properties, particularly those of interest to the telephone system. A compilation of representative data is given in Tables I and II.

The fundamental property which distinguishes a ferromagnetic material is that when it is subjected to a magnetic field it develops magnetic flux considerably larger than similarly attained in air. The magnetizing forces of interest in telephone apparatus range from less than 10^{-3} to upwards of 10^3 oersteds, and the flux densities from less than 1 to 30,000 gausses or more. The relation of flux density B to magnetizing force H for important materials as first magnetized is given on a logarithmic scale in Fig. 1. The ratio of B to H is the permeability, which can be read on the diagonal scale [8] in the figure. It is evident that the initial permeability μ_0 and the maximum permeability μ_m vary over a wide range from the hard magnet steels to the softest magnetic alloys. For commercial materials, 4–79 Mo-permalloy gives the largest initial permeability—around 22,000, and 78.5 permalloy gives the largest maximum permeability—about 105,000.

[5] G. W. Elmen, *Jour. Frank. Inst. 207*, 583 (1929).
[6] O. E. Buckley, *Jour. Applied Phys. 8*, 40 (1937).
[7] G. W. Elmen, *Elec. Engg. 54*, 1292 (1935).
[8] Scale due to Aiken; *Jour. Applied Phys. 8*, 470 (1937).

TABLE I

SOFT MAGNETIC MATERIALS IN SOLID OR SHEET FORM

Per Cent Composition						Raw Cost ¢/lb.	Typical Anneal	Material	Initial Permeability	Maximum Permeability	Saturation	Hysteresis Loss at Saturation	Residual	Coercive Force	Resistivity	Curie Temperature	Hysteresis Coefficient	Density
Other	Mn	Mo	Ni	Co	Fe				μ_0	μ_m	$4\pi I_\infty$	W_∞	B_r	H_c	ρ	θ	$a \times 10^6$	σ
											gauss	erg/cm.³	gauss	oersted	microhm-centimeter			gm./cm.³
3 C, 2 Si					95	<1	800° (Centigrade)	Cast Iron	—	600	—	20,000	5,300	4.6	30	—	—	—
					99.94	7*	900° Pot	Magnetic Iron	250	5,500	21,500	5,000	13,000	1.0	10	770° C.	50	7.88
					99.98	7	1480° H_2+880° H_2	Magnetic Iron H_2 Purified	25,000	275,000	21,500	300	13,600	0.05	10	770	—	7.88
0.5 Si					99.5	7*	800° Pot	0.5 Si-Iron (Field)	250	3,700	21,000	4,500	12,800	0.8	18	760	—	7.7
4 Si					96	8*	800° Pot	4 Si-Iron (Transformer)	400	6,700	20,000	3,500	12,000	0.5	60	690	120	7.5
9½ Si, 5½ Al					85	3	Cast	Sendust	30,000	120,000	10,000	100	5,000	0.05	80	—	—	7.1
0.2 Cu			99.0		0.4	35	1000° Pot	Nickel	110	600	6,100	3,000	3,600	3.4	8	360	100	8.85
	0.6		45		54.4	17	1100° Pot	45 Permalloy	2,700	23,000	16,500	1,200	8,000	0.3	45	440	0.4	8.17
			50		50	18	Long 1200° H_2	Hipernik	3,000	70,000	16,500	220	7,300	0.04	35	500	—	8.25
	0.6		78.5		20.9	28	1000° +Quench	78.5 Permalloy	9,000	105,000	10,700	200	6,000	0.05	16	580	0.2	8.60
5 Cu	1.0		74		20	27	900° +Quench	Mumetal	7,000	80,000	8,500	200	6,000	0.05	25	—	—	8.58
3.8 Cr	0.6		78.5		17.1	29	1000° Pot	3.8–78.5 Cr-Permalloy	10,000	40,000	8,000	200	4,500	0.05	65	455	0.3	8.56
	0.6	4	79		16.4	32	1000° Pot	4–79 Mo-Permalloy	22,000	72,000	8,500	200	5,000	0.05	55	460	0.05	8.72
15 Cu	1.0	3	71		10	29	1100° H_2	1040 Alloy	40,000	100,000	6,000	200	2,500	0.014	55	290	—	8.76
		12.5	80		7.5	40	800° H_2	12.5–80 Mo-Permalloy	9,000→1, over room temperature range					—	—	40	—	8.9
				99		136	1000° Pot	Cobalt	70	240	18,000	2,000	5,000	10	9	1120	30	8.9
				50	50	69	900° Pot	Permendur	800	5,000	24,500	12,000	14,000	2.0	7	1000	—	8.3
2 V				49	49	73	800° Pot	2 V-Permendur	800	4,500	24,000	6,000	14,000	2.0	26	980	1.0	8.2
	0.6		45	25	29.4	50	1000° +425° Bake	45–25 Perminvar	365	1,800	15,500	4,000	3,300	1.4	19	715	0.01	8.6
	0.6		70	7	22.4	35	1000° +425° Bake	7–70 Perminvar	850	4,000	12,500	—	2,400	0.6	16	650	0.06	8.6
	0.6	7.5	45	25	21.9	57	1000° +425° Bake	7.5–45–25 Mo-Perminvar	550	3,800	10,300	2,600	4,300	0.6	80	540	0.1	8.66

* Approximate finished cost for 14 mil sheet.

Note: Values for μ_0, μ_m, W_∞, B_r, H_c and a are subject to considerable variations, depending on purity of materials, composition, type of heat treatment, and final condition of mechanical stress.

TABLE II

PERMANENT MAGNET STEELS

Per Cent Composition					Raw Cost ¢/lb.	Typical Heat Treatment		Material	Initial Permeability	Reversible Permeability	Saturation	Residual	Coercive Force	Energy Product	B for (B·H) =Maximum	Resistivity	Curie Temperature	Density
Other	Cr	Ni	Co	Fe		Quench	Aging		μ_0	μ_r	$4\pi I_\infty$	B_r	H_c	$(B\cdot H)_{max.}$		ρ	θ	σ
											gauss	*gauss*	*oersted*		*gauss*	*microhm-centimeter*		*gm./cm.*^{3}
0.6 C, 0.8 Mn				98.8	0.8	800° C. Water	—	Manganese	75	—	21,000	10,000	50	0.2×10^6	6,900	20	750° C.	7.8
0.6 C, 0.4 Mn	1			98	0.9	800° Oil	—	1% Chrome	—	—	—	9,800	50	0.2	6,900	—	—	—
1 C, 0.4 Mn	3			96	1.1	840° Oil	—	3% Chrome	—	31	—	9,700	65	0.3	6,100	38	—	7.7
5 W, 1 C				94	9	840° H_2O	—	5% Tungsten	—	32	—	10,800	60	0.3	7,000	30	—	8.0
7 W, 0.5 Mn	3.5		36	52	62	940° Oil	—	Cobalt (Honda KS)	7	9.4	19,000	9,500	220	0.9	6,000	27	700	8.3
6.7 Ti, 3.7 Al		18	27	45	53	Cast	650°	Honda, New	3	3.8	—	7,100	780	2.0	4,100	65	—	7.3
13 Al		29		58	14	1200° Oil	600°	Mishima	4	5	11,600	6,000	550	1.4	3,500	60	750	7.1
14 Al		25		60	12	1200° Oil	600°	Oerstit 500	—	5	—	6,500	500	1.4	4,400	60	—	—
12 Al		20	5	63	17	1200° Oil	600°	Alnico I	—	4	—	7,300	430	1.4	4,500	60	—	7.0
10 Al, 6 Cu		17	12.5	54.5	26	1200° Oil	600°	Alnico II	—	4	—	7,200	540	1.6	4,400	—	—	7.1
17 Mo			12	71	33	1300° Oil	700°	Remalloy	8	12	17,000	10,500	250	1.1	6,500	45	780	8.4
60 Cu		20		20	13	1000° Oil	600°	Magnetoflex[9]	3	3	5,000	3,400	390	0.5	1,800	—	—	—
41 Cu		24	35		60	1050° Oil	600°	Magnetoflex[10]	4	4	8,600	5,300	440	1.0	3,400	38	850	8.7
77 Pt			23		$400	1200° Oil		77 Platinum—Cobalt	1.1	1.1	—	4,500	2,600	3.8	2,500	50	—	—
2 Fe_2O_3+1 Fe_3O_4 +1 Co_2O_3					25¢	950° Vacuum + Magnetize at 500°		Oxide, Compressed	1.7	1.7	—	1,800	600	—	—	10^{12}	350	3.8

[9] Neumann, Bücher & Reinboth, *Z. f. Metallkunde 29*, 173 (1937).
[10] Dannöhl & Neumann, *Zeit. f. Metallkunde 30*, 217 (1938).

Note: Values for μ_0, μ_r, B_r, H_c and $(B\cdot H)_{max.}$ are subject to considerable variations, depending on purity of materials, composition, and heat treatment.

TABLE III

Curve Number	Material	Typical Heat Treatment (Temperature in Degrees C.)	Initial Permeability μ_0	Maximum Permeability μ_m	Saturation $4\pi I_\infty$	Hysteresis Loss at Saturation W_∞	Residual B_r	Coercive Force H_c	Resistivity ρ
	Soft Magnetic Materials				*gauss*	*erg/cm.*^{3}	*gauss*	*oersted*	*micro-ohm-cm.*
1	Magnetic Iron	900 Anneal	250	5,500	21,500	5,000	13,000	1.0	10
2	4 Per Cent Silicon Iron	800 Anneal	400	6,700	20,000	3,500	12,000	0.5	60
3	45 Permalloy (Ni, Fe)	1100 Anneal	2,700	23,000	16,500	1,200	8,000	0.3	45
4	78.5 Permalloy (Ni, Fe)	1000 + Air Quench	9,000	105,000	10,700	200	6,000	0.05	16
5	4–79 Mo-Permalloy (Mo, Ni, Fe)	1000 Anneal	22,000	72,000	8,500	200	5,000	0.05	55
6	2 V Permendur (V, Co, Fe)	800 Anneal	800	4,500	24,000	6,000	14,000	2.0	26
7	45–25 Perminvar (Co, Ni, Fe)	1000 + 425 Bake	365	1,800	15,500	4,000	3,300	1.4	19
	Magnet Steels					Energy Product $(B \cdot H)_m$ $\times 10^{-6}$			
8	3 Per Cent Chrome	840 Quench	10	100	—	0.34	9,700	65	38
9	Honda (C, W, Co, Fe)	940 Quench	7	—	19,000	0.9	9,500	220	27
10	Mishima (Al, Ni, Fe)	Quench + 600 Bake	4	16	11,600	1.4	6,000	550	60
11	Remalloy (Mo, Co, Fe)	Quench + 700 Bake	12	30	—	1.1	10,500	250	45
12	Oxide (Fe, Co, O_2)	950 Vacuum	1.7	—	—	—	1,800	600	10^{12}

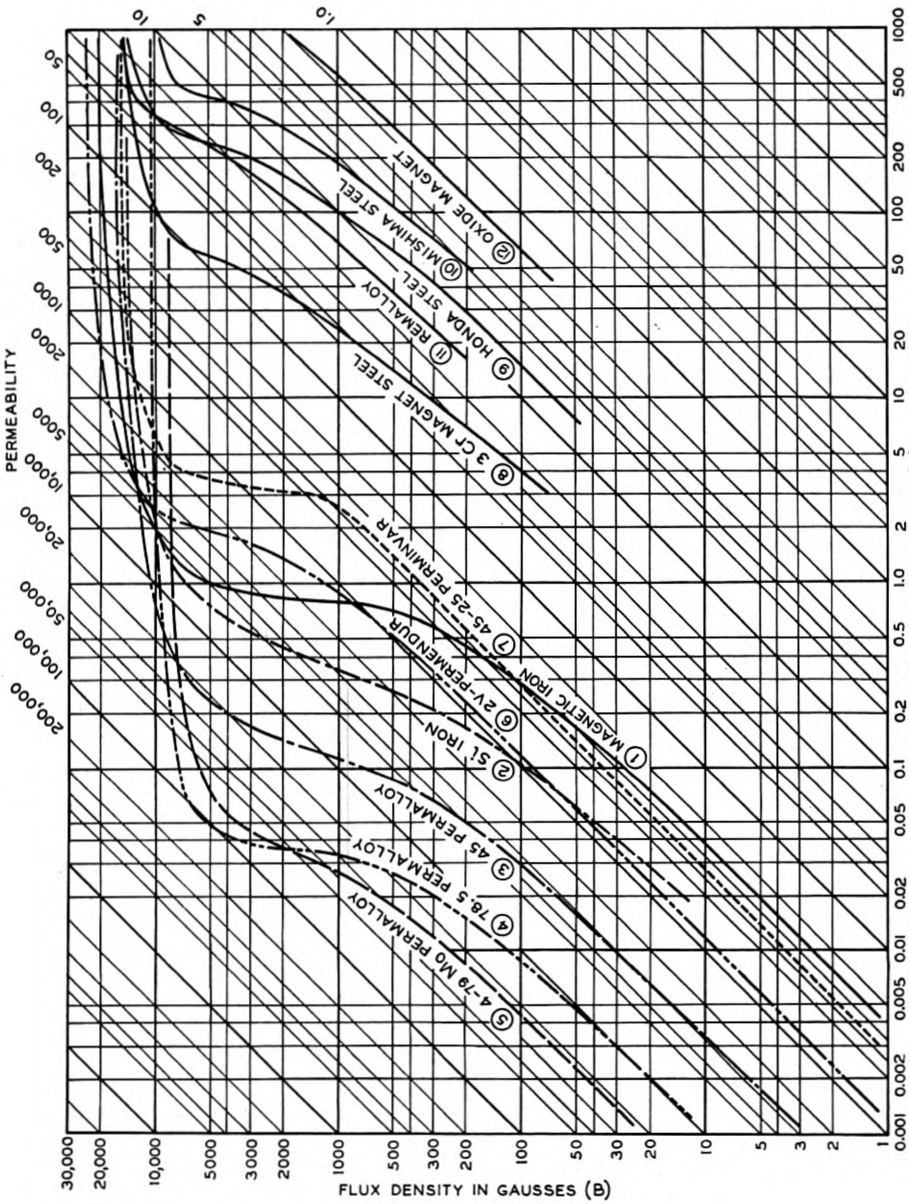


Fig. 1—Magnetization and permeability curves for important magnetic materials (diagonal scale for permeability values).

The saturation flux density for the soft magnetic materials can be read from these curves, and it is listed in the adjacent Table III in the column headed $4\pi I_\infty$. The pre-eminence of 2 V-permendur is noteworthy.

Further important magnetic properties are obtained from the hysteresis loop. This gives the B, H relationship for a material which has already been magnetized up to a peak value H_m. The flux density remaining after the removal of a very large magnetizing force is the residual B_r, and the reverse magnetizing force necessary to bring the flux to zero is the coercive force H_c. The area of the B, H loop when the peak magnetizing force is very large gives the maximum energy W_∞ dissipated by hysteresis in the material when it is carried from positive magnetic saturation to negative, and back again. Table III gives values of W_∞, B_r and H_c. The low values of these properties for 4–79 Mo-permalloy are noteworthy. Among soft magnetic materials, iron and 2 V-permendur have high residual and coercive force, properties which are occasionally useful.

Permanent magnet materials should have large values of B_r and H_c, although a sacrifice of residual can be more or less compensated by an increase in coercive force. A more fundamental criterion of permanent magnet quality is the peak energy product $(B \cdot H)_{max.}$ obtained from the demagnetizing section of the hysteresis loop.[11] Values of this product are given for several magnet steels in Table III. Mishima steel is seen to be foremost in this regard, with remalloy nearly as good.

The last column in Table III gives the resistivity ρ of each material. A high resistivity such as obtained with molybdenum additions to permalloy and perminvar is desirable in suppressing eddy current losses for alternating current applications. Eddy current losses can also be suppressed by proper subdivision of the material, but this method becomes costly with fine subdivision.

For apparatus depending upon the tractive force of a magnet, the high flux density properties of materials are most important. Since these do not show up clearly in Fig. 1, an accompanying Fig. 2 has been prepared in which the $(B - H)$ scale is quadratic, and thus proportional to tractive force. The relative merits of various materials for such applications are seen by inspection of these curves. 2 V-permendur is outstanding at high flux densities, iron and 45 permalloy at intermediate, and 4–79 Mo-permalloy at low flux densities.

The use of a purifying hydrogen anneal has been shown by Cioffi [12] to increase the ease of saturating iron and most magnetic alloys. The

[11] S. Evershed, *J. I. E. E.*, London, *58*, 780 (1920); *63*, 725 (1925).
[12] P. P. Cioffi, *Phys. Rev. 39*, 363 (1932).

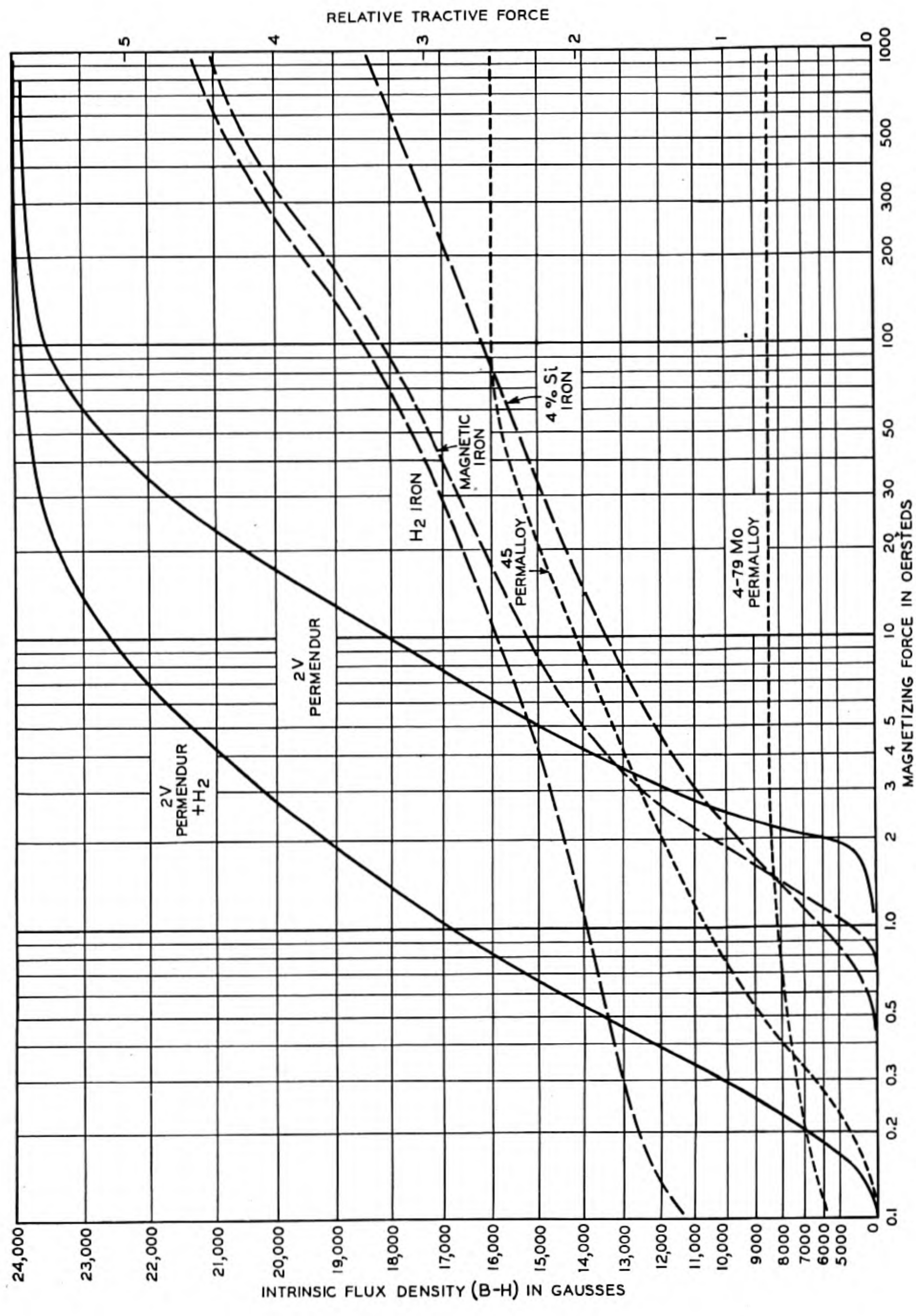


Fig. 2—Magnetization curves plotted to show the relative tractive force for important magnetic materials.

in the direction of rolling [19] (see Table IV). Such materials have been used to some extent abroad for loading coil cores. Aside from the difficulty of attaining as low hysteresis loss, high permeability, and high stability as obtainable with permalloy powder cores, such materials appear to be at an economic disadvantage because of the high cost of rolling sheet thin enough for eddy current loss reduction.

Stresses well within the elastic limit increase or decrease the permeability of materials, depending on the composition.[20] Furthermore, tension and compression generally have opposite effects. The changes in permeability due to such weak stresses are not permanent, but practically disappear when the stress is removed.

Cost of Magnetic Materials

Many materials having very desirable technical qualifications cannot be used extensively because of their high costs. Thus, nickel at 35¢/lb., cobalt at $1.35/lb., tungsten at $1.80/lb., and vanadium at

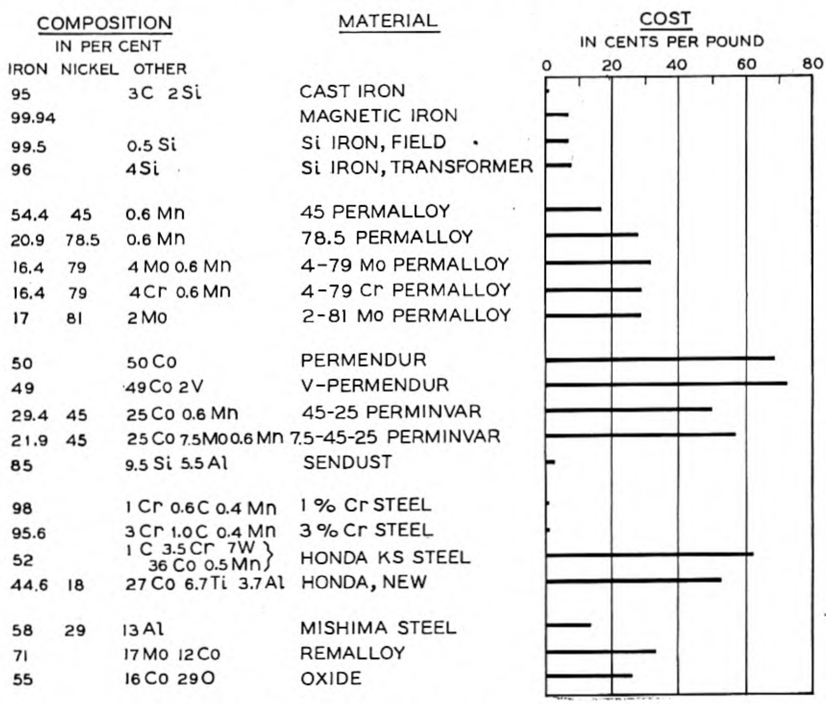


Fig. 3—Raw materials cost of various magnetic alloys.

[19] O. v. Auwers, *Wiss. Ver. a.d. Siemens-Werk. 15*, 112 (1936).
[20] O. E. Buckley & L. W. McKeehan, *Phys. Rev. 26*, 261 (1925).

\$2.80/lb. increase the cost of an alloy very considerably in comparison with iron or steel at less than 7¢/lb. High priced magnetic alloys can only be justified in general when their extraordinary characteristics permit offsetting apparatus performance or economies. Of course, they may be absolutely necessary for certain types of apparatus.

The cost data for Tables I and II have been calculated from recent prices [21] for raw materials of quality suitable for magnetic alloys. The cost of raw materials is given in Fig. 3 for selected alloys. The low raw materials costs of iron, silicon-iron, and chromium steel are notable, as well as that of the powder core material "Sendust." 45 permalloy is the cheapest of the high permeability materials, while Mishima steel is the cheapest of the high quality permanent magnet materials.

Comparisons based on raw materials costs are not entirely satisfactory. The cost of alloying and reducing to finished form may overshadow the cost of raw materials, particularly when high purity, exact tolerances, and small rates of production are involved.

Applications

Almost all magnetic properties are utilized in some type of telephone apparatus. They are generally linked inseparably with electrical and mechanical properties. The proper design of any apparatus strikes a compromise between the various technical features and cost. The technical features of materials used in present day apparatus are listed below. Acceptable common properties, such as mechanical soundness and workability, are assumed for all materials unless specifically mentioned. Listing is made on the basis of the magnetic effect utilized.

1. *Simple Tractive Force* (*Relays*)

The force of attraction between two neighboring surfaces of area A, between which the flux density is B, is

$$F = kAB^2.$$

The primary telephone application of this effect is to relays and switches. For greatest tractive force, materials capable of attaining high flux densities are desirable. However, the air gap in the magnetic circuit absorbs such a large proportion of the available magnetomotive force that higher or lower permeabilities in the core material are frequently less important than efficiency of design. A typical relay structure is shown in Fig. 4.

[21] *Steel*, Oct. 3, 1938.

Reference to Figs. 1 and 2 points to 2 V-permendur as outstanding in the flux density range above 15,000 gausses. This material is excluded for many applications because of its high cost. High temperature hydrogen annealing improves the high flux density behavior of most magnetic materials, as noted above in connection with Fig. 2. Using ordinary methods of annealing, the next best material for high flux operation is the standard magnetic iron, while 45 permalloy is preferable at flux densities below 12,000. For the low magnetizing forces available in sensitive relays, 4–79 Mo-permalloy gives the largest tractive forces.

There are frequently other requirements in addition to tractive force in relay construction. The operation and release characteristics

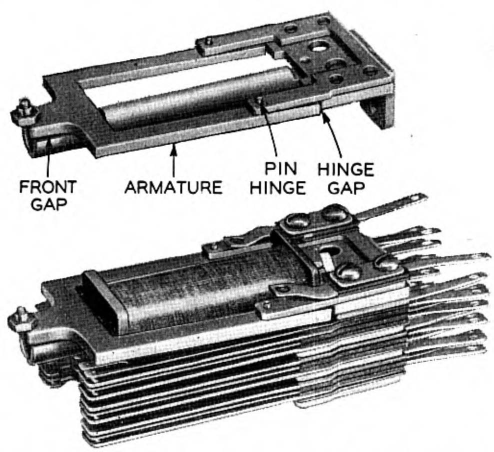


Fig. 4—U-type relay, showing the magnetic circuit

are determined by the resistivity and coercive force of the material. Sensitive, quick release relays require materials like permalloy, while slow release types may utilize the larger coercive force and residual of magnetic iron or cold rolled steel.[22, 23]

Relays for a-c. applications may have objectionable eddy current losses in their cores. Such losses can be reduced by use of high resistivity material. Thus, 45 permalloy has one-fourth the loss of iron for the same core thickness and flux density.

2. *Polarized Tractive Force* (*Receivers, Ringers, Relays*)

If permanent polarizing flux density B_p exists between two neighboring surfaces, and a small additional flux Φ is applied by means of a

[22] H. N. Wagar, *Bell Labs. Record 16*, 300 (1938).
[23] F. A. Zupa, *Bell Labs. Record 16*, 310 (1938).

magnetomotive force M, the additional tractive force will be approximately

$$F = 2kB_p\Phi = \frac{2kB_pM}{R_a(1 + n/\mu_r)},$$

where R_a in the latter part of the above equation is the reluctance of the air-gap, n is ratio of the reluctance of the ferromagnetic circuit with iron removed to the reluctance of the air-gap, and μ_r is the reversible permeability, i.e. the permeability measured with very small a-c. magnetizing forces in the presence of the polarizing flux density. It thus appears that materials for such applications should have high saturation values. The apparatus should be designed to obtain a low value of n, and to operate at a flux density to make the above force a maximum.

Figure 5 gives values of μ_r as a function of polarizing or superposed flux density B_p. It should be remarked that the reversible permeability is practically single valued [24] when plotted against polarizing flux, in contrast with the "butterfly" loop obtained by plotting against polarizing field strength. The superiority of permendur in Fig. 5 is obvious. Values of the force factor for $n = 100$ and $n = 1000$ have been computed for these materials for an arbitrary value of air-gap reluctance and magnetomotive force. It is evident that the full advantage of permendur is not realized unless the apparatus design is such as to attain a low value of the air reluctance ratio n.

Permanent magnet materials are frequently employed to supply polarizing flux. When they form a part of the circuit for the alternating flux, their reversible permeability becomes important. Reference to Table II shows that 5 per cent W steel has the highest permeability of the common permanent magnet materials. However, its energy product $(B \cdot H)_{max}$ is so low that other materials are preferred for applications where space is limited, despite their low reversible permeabilities.

The earliest application of polarized structures in the telephone plant was to the receiver. The receiver of the present day is constructed with remalloy permanent magnets, 45 permalloy pole-pieces, and a permendur diaphragm.[25] The magnetic circuit is shown in Fig. 6.

The second application of polarized structures is to the telephone bell or ringer. Here a permanent magnet is used, to supply polarizing flux to the two cores for the coils through the shoe at one end and the

[24] R. Gans, *Phys. Zeit. 12*, 1053 (1911).
[25] W. C. Jones, *B. S. T. J. 17*, 338 (1938).

armature at the other.[26] Non-magnetic stop pins are required on the armature to prevent sticking. This complicated magnetic circuit has been developed to meet the numerous requirements placed on ringers. Polarization is necessary for an a-c. ringer which is to operate without

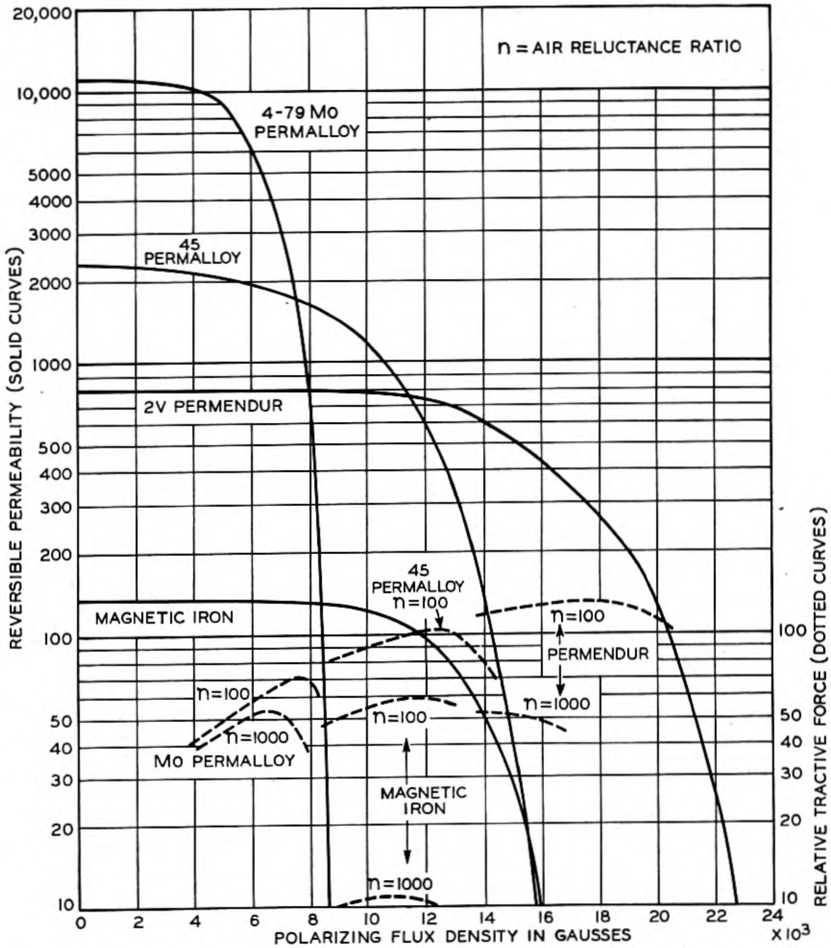


Fig. 5—Reversible permeability and relative polarized tractive force for various magnetic alloys.

interrupter contacts, and it is employed in selective ringing on party lines. A high coil inductance is required to limit the transmission

[26] K. B. Miller, "Manual Switching and Substation Equipment" (McGraw-Hill, 1933 ed.), p. 67.

losses due to the shunting effect of ringers across the line, especially in the case of party lines.

The third application of polarized structures is in certain relays for composite ringing, duplex telegraph, and special selecting circuits. Various materials are used in such relays, depending upon the sensi-

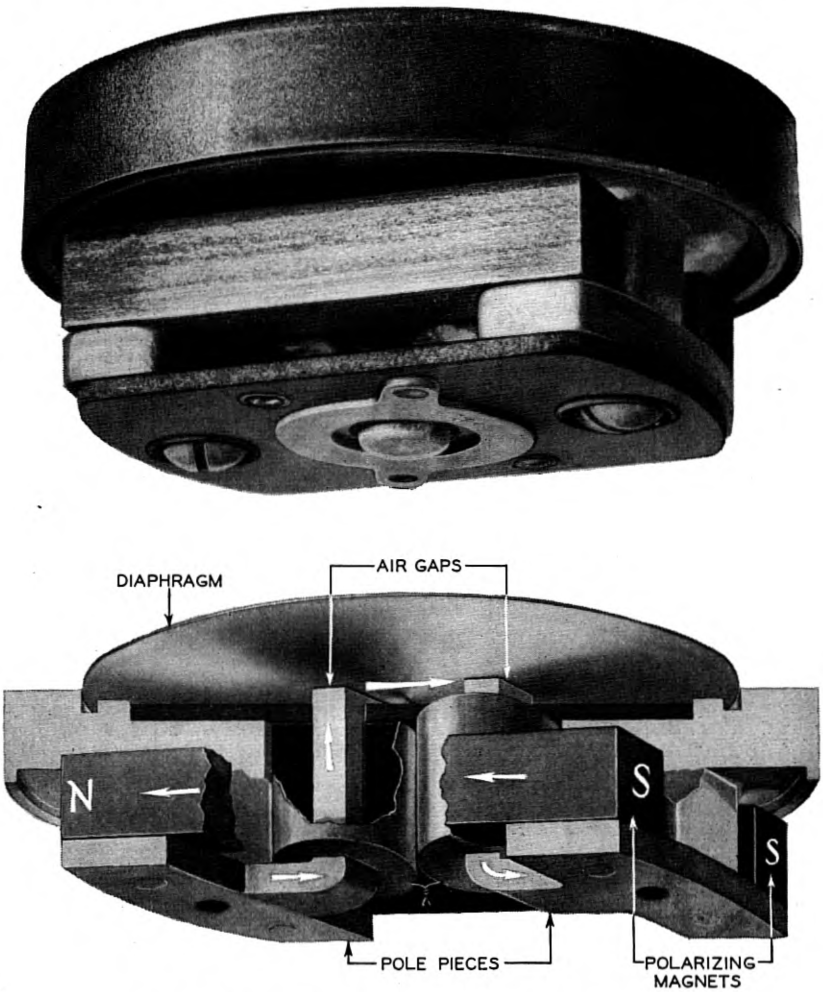


Fig. 6—Magnetic circuit of the new telephone receiver.

tivity required, so that no general statement can be made as to needs in this field. The problems encountered are essentially similar to those met in the receiver and ringer, for which numerous materials are available.

3. *Force on Current* (*Moving Coil Receivers, Light Valves, Motors*)

A straight wire of length l carrying a current i in a perpendicular magnetic field of flux density B is pushed at right angles to the field and the length with a force

$$F = Bil = \mu Hil,$$

where μ is the permeability and H is the magnetizing force in the nearby material from which the flux is derived. Again, the prime requirement for a useful material is high flux capacity, and high permeability, so that the magnetizing force need not be large. The magnetizing force has been supplied generally in the past by means of direct current in windings built into the apparatus. With the normally available voltages and currents, sufficient magnetizing forces could be obtained only with coils having a large number of turns, and low resistance. This generally involved such large structures that cost considerations compelled the use of iron cores with perhaps pole-pieces made of permendur. Lately many structures are being designed to replace costly electromagnets with permanent magnets made from Mishima type steel.

Moving coil receivers and loud speakers [27] are the most important representatives of this type of apparatus. Others are the string oscillograph,[28] the light valve,[29] phonograph record recorder,[30] and various types of power machinery.[31] Several of these are now constructed with permendur pole-pieces, and cast Mishima steel magnets, or remalloy magnets where hot rolling will assist in producing small, accurately sized parts.

4. *Induced Electromotive Force*

The electromotive force between the terminals of a coil of N turns linking flux φ is

$$- e = N \frac{d\varphi}{dt}.$$

The arrangement of coils and interlinking flux differs considerably in the various types of apparatus employing this effect.

The flux variation is provided by means of mechanical motion of the coil in instruments such as the electromagnetic microphone. It is varied by means of fluctuations in magnetizing current in inductance

[27] E. C. Wente & A. L. Thuras, *B. S. T. J. 10*, 565 (1931).
[28] A. M. Curtis, *B. S. T. J. 12*, 76 (1933).
[29] G. E. Perreault, *Bell Labs. Record 10*, 412 (1932).
[30] H. A. Frederick, *B. S. T. J. 8*, 159 (1929).
[31] R. D. deKay, *Bell Labs. Record 16*, 236 (1938).

coils and transformers. In the moving coil microphone, the cylindrical coil moves axially in a slot between an inner magnetized cylinder and an outer magnetic cylinder which receives the radial flux threading the coil. For such a cylindrical coil in a uniform radial field, the e.m.f. is $-e = lB\frac{dx}{dt}$, where l is the total length of wire composing the coil, and x is its displacement perpendicular to the radial field. For a stationary coil linking a varying flux as in a transformer, $-e = AN\frac{dB}{dt} = AN\mu\frac{dH}{dt}$, where A and μ are the area and permeability of the core within which the magnetizing force is H. It is evident from these equations that high flux density or permeability are desirable, in order to yield the largest e.m.f. with least material.

4a. *Microphones, Magnetic Tape Recorders, Magnetos*

An application involving a moving coil is the public address microphone,[32] where the flux is established by means of a cobalt steel magnet, and is concentrated upon the moving coil by means of permendur pole-pieces.

An inverse application in which the coil is stationary and the magnet moves is the magnetic tape recorder.[33] In this, a steel tape which has been magnetized by speech currents is drawn between permalloy pole-pieces in pick-up coils. High coercive force, high signal-to-noise ratio, mechanical soundness, durability, and cheapness of the tape material are desirable.

The telephone magneto employs the e.m.f. generated by rotating a coil in a magnetic field. It has been constructed with iron armature and pole-pieces, and chrome steel field magnets. Recent designs using modern magnetic materials have indicated the possibility of large economies in volume.[34] The magnetic properties of available materials have now reduced the volume required by magnetic parts to a point where the major problem in magneto design is to compress the gears and shafts into correspondingly small space and yet maintain sufficient mechanical strength, durability, and convenience of operation.

4b. *Inductance Coils*

A very important application of induced e.m.f. focuses attention on the inductance of a coil of N turns surrounding a (closed) core of area A,

[32] R. N. Marshall & F. F. Romanow, *B. S. T. J. 15*, 405 (1936).
[33] C. N. Hickman, *B. S. T. J. 16*, 165 (1937).
[34] *Ericsson Bulletin No. 12*, 46 (1938).

magnetic path length l, and permeability μ. The inductance is increased through the presence of the core by an amount

$$L = 4\pi N^2 \mu A / l.$$

As noted earlier, when eddy current shielding is negligible such an inductance is accompanied at frequency f by hysteresis, residual, and eddy current resistances to give a total as follows:

$$R/L = \mu f(aB_m + c + ef).$$

At frequencies high enough to introduce eddy current shielding, the effective inductance due to a core of laminar thickness t and resistivity ρ is reduced below the ordinary inductance L_0 by the ratio

$$L/L_0 = \frac{1}{\theta} \frac{\sinh \theta + \sin \theta}{\cosh \theta + \cos \theta},$$

where $\theta = 2\pi t \sqrt{\mu_0 f / \rho}$. Figure 7 shows this ratio and the ratio $R_e/\omega L_0$ as functions of θ. As a practical example, the permeability of 6 mil (0.015 cm.) 4–79 Mo-permalloy is reduced to about 75 per cent of its initial value (22,000) at 1 kilocycle, and to about 17 per cent at 10 kc.

The telephone loading coil adds inductance to the telephone line, but it must not add excessive resistance. Furthermore, its inductance must be extremely stable with the lapse of time, and under severe operating conditions, such as occasional current surges induced from lightning discharges. Iron wire cores for loading coils were supplanted over twenty years ago by compressed iron powder cores, and these in turn gave way to permalloy powder cores. The latest improvement is the introduction of 2–81 Mo-permalloy powder cores.[7] The reduction in size of cores with these improvements is shown in Fig. 8.

Another method of loading a line is by sheathing the conductor with a continuous layer of magnetic material. This method was used with notable success on submarine telegraph cables by wrapping permalloy tape upon the conductor, and annealing before applying insulation.[3] The location of the loading material is shown in Fig. 9. The continuous loading of long submarine telephone cables has been shown to be feasible using thin 7.5–45 Mo-perminvar tape.[7]

Retardation and choke coils have a great variety of applications, running from a tiny coil weighing $3\frac{1}{2}$ ounces [35] to a 4600 lb. generator ripple suppressor.[33] The contrast is evident in Fig. 10. Retardation

[35] D. W. Grant, *Bell Labs. Record 11*, 173 (1933).
[36] R. A. Shetzline, *Bell. Labs. Record 17*, 34 (1938).

coils, used as network elements, are generally equipped with compressed powder cores, and the same improvements are indicated as for loading coils noted above. Laminated permalloy cores are often used for low-frequency applications (ringing, telegraph), where high inductance is desired, and a-c. losses are naturally low.

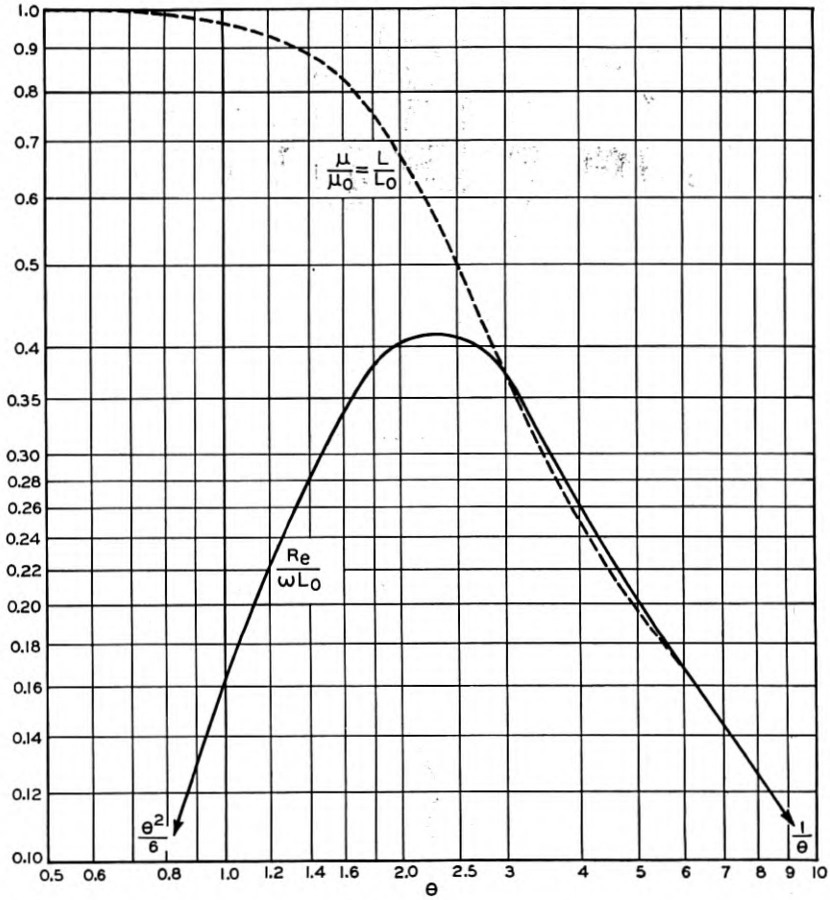


Fig. 7—Effect of eddy current shielding on apparent permeability and on eddy current resistance/reactance ratio of sheet core material.

Coils designed for direct currents must have cores with high a-c. permeability in the presence of superposed field, i.e. they must be made of high permeability magnetic materials having high saturation values. Reference to Fig. 5 shows permendur outstanding in this regard. In practice, the low costs of silicon iron or magnetic iron are frequently decisive in the selection of these materials instead of the technically superior materials noted above. Air-gaps are often found necessary to reduce the superposed field strength in the magnetic core.

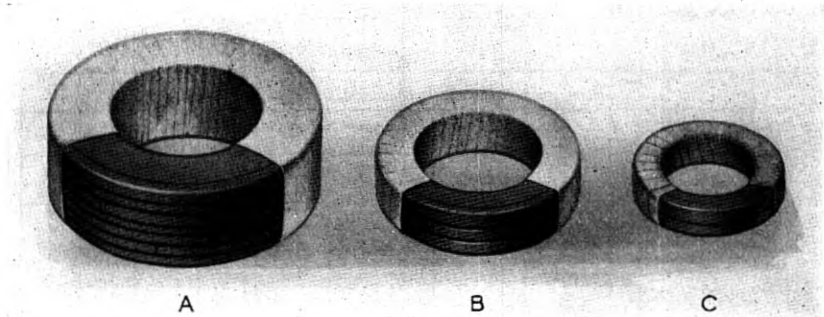


Fig. 8—Relative sizes of compressed powder cores for loading coils; *A*. Iron, *B*. 80-permalloy, *C*. 2–81 Mo-permalloy.

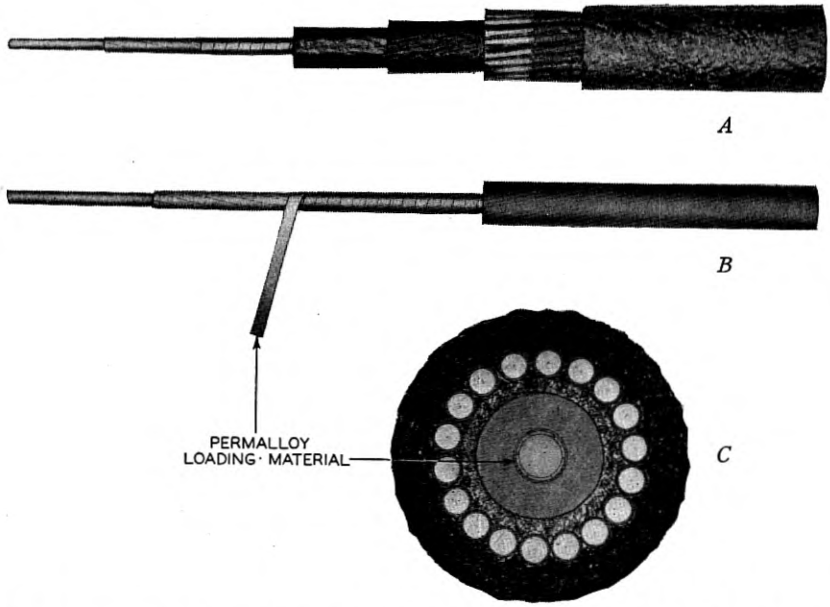


Fig. 9—Permalloy loaded telegraph cable; *A*. Armored deep sea type, *B*. Gutta percha insulated core, *C*. Enlarged cross sectional view.

This serves to retain a fairly large reversible permeability in the core, and, if the air gaps are not too large, it yields a larger net effective permeability than could be obtained without air-gaps.

A further type of inductance is the impulse coil used in harmonic generators.[37] The cores of these coils should be saturated over most of the magnetizing cycle, and reverse very quickly and completely just as the magnetizing force passes through zero. This implies use of material having a high permeability, and a high resistivity, such as 4–79 Mo-permalloy.

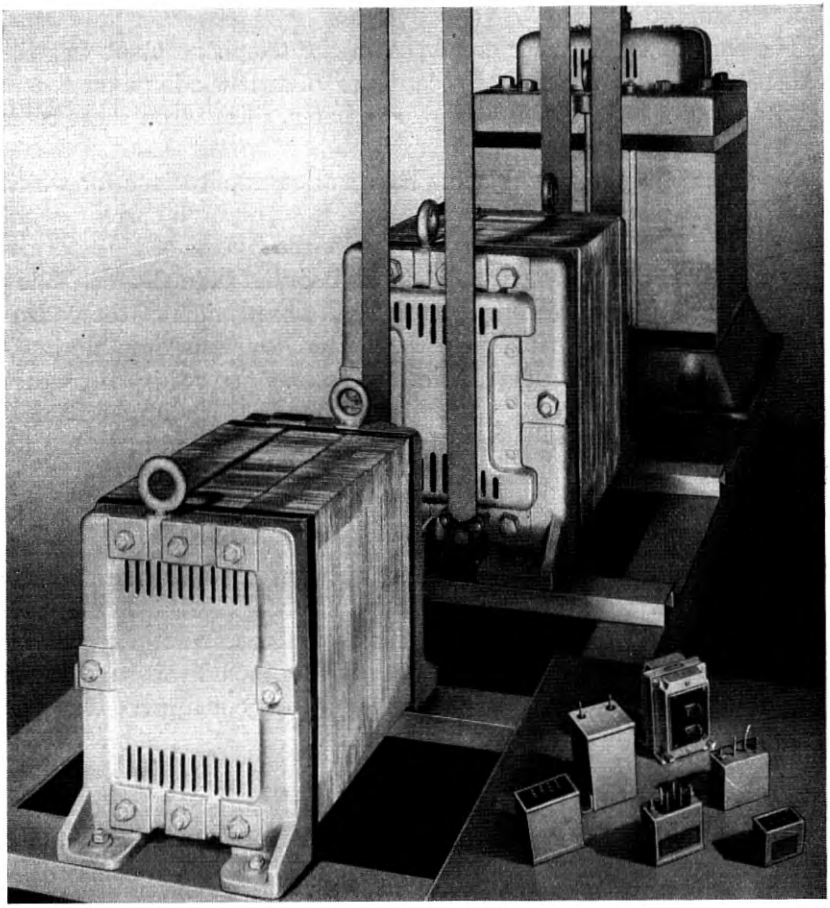


Fig. 10—Large and small coils used in the telephone system.

4c. *Transformers*

With transformers, the inductance and losses of the individual coils can be analyzed as simple inductances, for which the considerations of

[37] E. Peterson, J. M. Manley & L. R. Wrathall, *Elec. Engg.* *56*, 995 (1937).

4b apply. In addition, however, the coupling factor between coils on the same core becomes of especial importance. In the usual design the flux linkage common to the primary and secondary is largely contained in the magnetic material, while the leakage flux is controlled by the reluctance of the air path. It is thus evident that a large value of core permeability is required to obtain a high coupling factor. Of course, advantages indicated by high permeability may be lost through improper design.

The earliest transformer employed in the telephone plant was the induction coil. This originally consisted of two windings on a core composed of a bundle of iron wires.[38] Later, silicon iron sheet cores were introduced.

Input and output transformers have varied applications for which special types of cores are required. Where superposed current is not involved, space and weight can be economized by use of high permeability materials such as chrome or molybdenum permalloy.[39] Eddy current shielding at higher frequencies will offset much of the permeability advantage indicated for these materials unless they are laminated sufficiently. However, thin laminations are costly to prepare and stack, and difficult to insulate and handle without mechanical injury and corresponding reduction of permeability. An intermediate thickness of permalloy sheet is generally chosen, which secures a considerable advantage in permeability over iron, without prohibitive cost.

Where a winding must carry direct current, conditions are similar to those applying to choke coils, and materials with high reversible permeability at high magnetizing forces are required. Frequently, for large magnetizing forces, it is desirable to include air-gaps in the magnetic circuit. Silicon iron is the ordinary material for such application, as it is for power transformers.

4d. *Magnetic Shielding*

A further application of induction effects is in magnetic shielding of apparatus. A magnetic shield consists of a high permeability shell (4–79 Mo-permalloy, or 78.5 permalloy) which shunts flux around the enclosed apparatus. For a-c. shielding, alternate layers of copper and permalloy sheet are very effective in magnetic shunting and eddy current screening of the enclosed space.[40] High initial permeability,

[38] Cf. p. 43 of Reference 26.
[39] A. G. Ganz & A. G. Laird, *Elec. Engg. 54*, 1367 (1935).
[40] W. G. Gustafson, *B. S. T. J. 17*, 416 (1938).

mechanical workability, and low cost are desirable for such applications.

5. *Magnetostriction*

The relative change in length of a magnetic bar upon magnetization, $\delta l/l$, ranges from negative to positive values, depending on alloy composition, and is roughly proportional to B^2. If a polarizing field is applied, small additional alternations of field will give accompanying and nearly proportional alternations of length of a bar. These alternations evidence themselves in the electrical constants of a coil enclosing the bar.

This effect can be utilized in oscillators and filters for frequencies which involve the use of mechanically resonating bars of convenient size. Among the high permeability materials, 45 permalloy appears to give the largest magnetostrictional effect.[41] In order to limit eddy current losses, the material must be laminated more or less finely, depending on the frequency.

The inverse magnetostriction effect by which e.m.f. is generated in a coil when the core is vibrated, becomes objectionable as a source of circuit noise in the transformers of high gain amplifiers.[39] An alloy with minimum magnetostriction such as 81 permalloy or 4–79 Mo-permalloy is preferred for such cases.

Other effects of magnetostriction are the generation of sound by the cores of coils subject to alternating magnetization, and the appearance of undesired resonance effects in the electrical circuit at frequencies at which the core resonates mechanically.

6. *Thermal Variation of Permeability*

The initial permeability of ordinary magnetic materials increases more or less slowly with increasing temperature, until a maximum value is reached, above which temperature the permeability declines very rapidly to the non-magnetic, or Curie point. The Curie point of an alloy can be moved down the temperature scale by adding non-magnetic materials, such as Mo, Cr, Cu, etc., to the alloy.[42]

The positive temperature coefficient of inductance of powder core coils due to thermal change of permeability becomes objectionable in crystal filters, where only very small variations in the resonant frequencies can be tolerated.[43] It is made slightly negative to counteract the small positive coefficient of mica condensers by the admixture with

[41] A. Schulze, *Zeit. f. Phys. 50*, 448 (1928).
[42] G. W. Elmen, *Bell Labs. Record 10*, 2 (1931).
[43] C. E. Lane, *B. S. T. J. 17*, 125 (1938).

the regular 2–81 Mo-permalloy powder of a small amount of permalloy powder containing about 12.5 per cent of molybdenum. The latter material has a Curie point just above room temperature.

Conclusion: Desirable and Possible Improvements

It appears from the above inspection of the magnetic elements of apparatus that there is a general desire for all properties which contribute to magnetic effects to be increased, and for those which cause apparatus energy losses to be decreased. A considerable success has already been achieved in these directions. The obstacles to further improvement are in part difficulties in commercial application of laboratory techniques, and in part ultimate limitations to the properties of materials.

The chart in Fig. 11 shows the recent trend toward the realization of the best possible magnetic properties in new materials or by improved processes. Values are given for the year 1920 (i.e. before commercial application of the permalloys), for 1939 as commercially and as experimentally realized, and for what may be considered as the attainable limit. It is evident that most of the properties which could be improved have been improved in commercial materials by a factor of ten or so within the last twenty years. A further improvement of several properties by a factor as large as ten has been observed by experimental procedures. These improvements have not been utilized in all cases, either because they may not be of great practical value, or because they involve processes which are commercially impracticable, or materials which are very expensive. Thus, the highest value of μ_m has been attained on a single crystal of H_2 purified iron [44] cut so as to make a hollow parallelogram with sides parallel to the (100) crystal axes, and annealed in H_2 below the α-γ transformation point. The highest permanent magnet quality has been attained with a platinum cobalt alloy [45] costing some $400 per pound.

One of the main objectives of commercial magnetics research has been the attainment of higher permeabilities—initial, maximum, reversible, and at high flux densities. The reversible permeability is closely linked with the initial permeability and saturation flux density. The permeability at high flux densities [46] appears to be susceptible to considerable increases by proper treatment of materials. The initial permeability can be increased by proper technique, but such gains are frequently sacrificed in practice because of unavoidable mechanical

[44] P. P. Cioffi & O. L. Boothby, *Phys. Rev. 55*, 673 (1939).
[45] W. Jellinghaus, *Zeit. Tech. Phys. 17*, 33 (1936).
[46] For example, at $B = 10{,}000$, or $B = 20{,}000$.

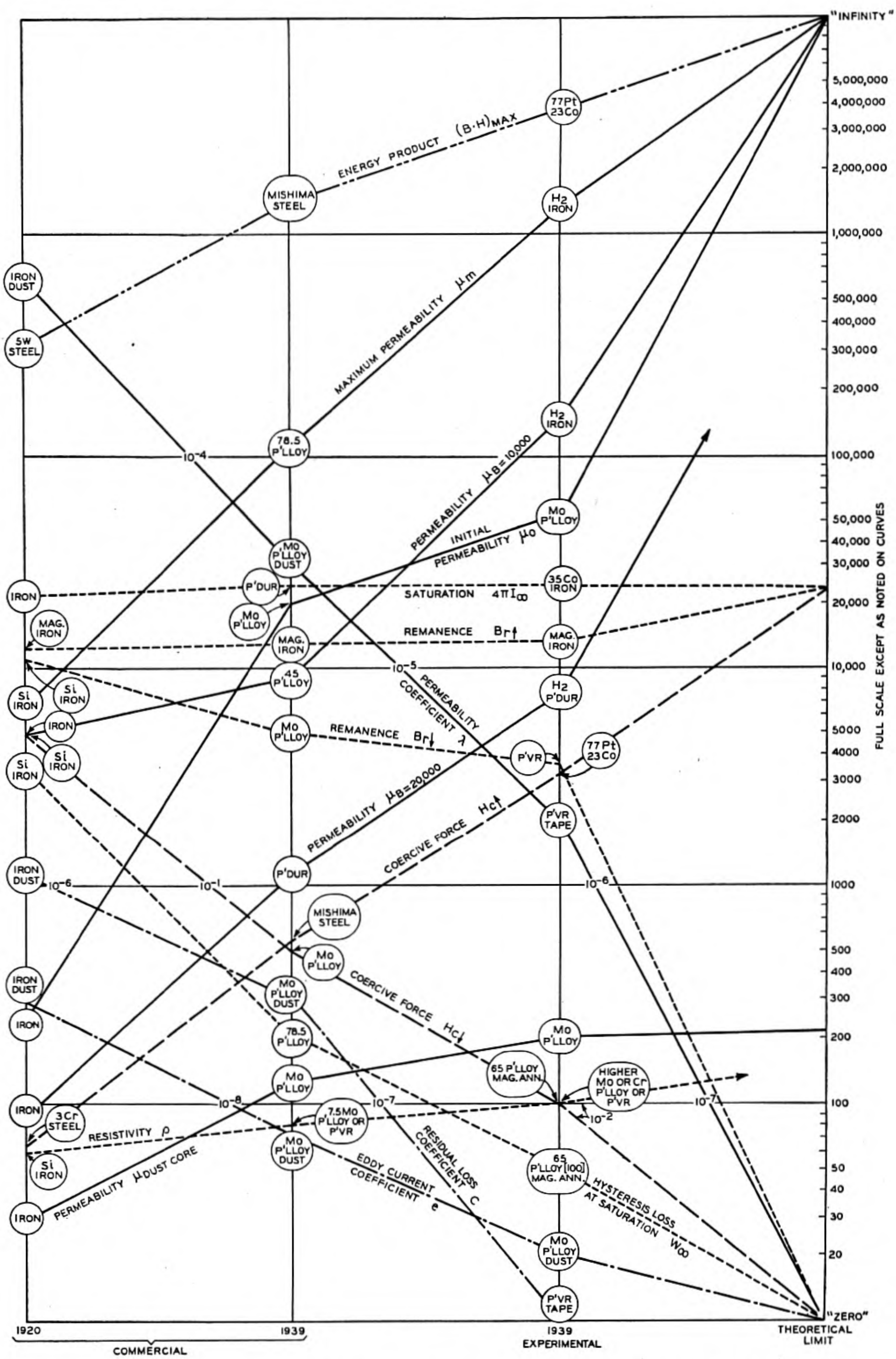


Fig. 11—Improvements in the properties of magnetic materials since 1920, in relation to theoretically possible properties.

stresses, or because of large eddy current shielding. Higher maximum permeabilities than now attainable do not promise great utility. However, the low values of coercive force and hysteresis loss found with materials having high maximum permeability may be sufficiently desirable, regardless of permeability needs.

Another important objective of magnetics research has been to reduce energy losses. Hysteresis loss at low flux densities, as indicated by the loop area coefficient a, should be decreased to cut down harmonic generation and modulation in magnetic core coils. Perminvar has shown desirably reduced losses, but it is sensitive to magnetic and mechanical conditions. Eddy currents are controlled by resistivity and degree of subdivision of the magnetic core. The resistivity of magnetic alloys can be increased to around 100 microhm-cms. by alloying with large enough quantities of chromium, molybdenum, etc., but at a serious sacrifice of magnetic quality for resistivities above about 60. Eddy current suppression by laminating or pulverizing the magnetic material thus offers a greater range of control than resistivity adjustment.

Permanent magnetic materials have also reached a very successful stage, from the magnetic point of view. The greatest handicap of the better materials is extreme hardness, which hampers fabricating processes.

Impedance Properties of Electron Streams

By LISS C. PETERSON

The input impedance of an idealized space charge grid tube is investigated under general conditions of space charge between the accelerating grid and the negatively biased control grid. It is shown that under certain space charge conditions the input capacitance and conductance both may be negative. These impedance properties persist up to frequencies for which the transit angle is quite large. Possibilities of designing electronic negative capacitances are thus opened up. Experimental results are also given; these give a broad confirmation of the theoretical deductions.

PART I

Theory

IN the early stages of vacuum tube history the theoretical work on d-c. space charge treated mainly potential distributions associated with fairly small initial velocities of the electrons. With the advent of multi-electrode tubes this situation changed for it then became necessary to consider also potential distributions occurring when electrons are injected with large initial velocities. Idealizations were introduced to the extent that consideration was given only to space charge conditions which can exist between two parallel planes at known potentials when electrons with normal velocities corresponding to these potentials are injected into the region through one or both planes.[1]

Some time ago it was discovered experimentally that space charge may under certain conditions produce a negative capacitance. The negative capacitances were found during a series of low-frequency measurements of the control-grid-to-ground capacitance of an experimental space-charge-grid tube. In the course of these measurements it was found that with all the electrodes carefully by-passed to ground for a-c. except the negatively polarized control grid, the input capacitance as well as the input conductance was negative in certain domains which depended upon the d-c. operating voltages.

In order to arrive at an understanding of this fact, a-c. phenomena must be considered under the general d-c. space charge conditions

[1] Plato, Kleen, Rothe, *Zeitschrift f. Phys.*, *101*, July 1936. C. E. Fay, A. L. Samuel, W. Shockley, *Bell Sys. Tech. Jour.*, January 1938. B. Salzberg, A. V. Haeff, *R.C.A. Review*, January 1938.

referred to. The investigation comprises the calculation of the impedance between two parallel planes in vacuum, at known d-c. potentials, when an electron stream with normal d-c. velocities corresponding to these potentials is injected at right angles to the planes. As a further step, the effect of a negatively polarized grid interposed between the two planes will be considered. This latter arrangement may be taken to correspond to an idealized space charge grid tube.

Before these calculations are presented it is well to set forth as concisely as possible those results of the d-c. space charge analysis [1] which will be of most immediate interest. For this purpose we start with a qualitative review of the work of Fay, Samuel and Shockley.

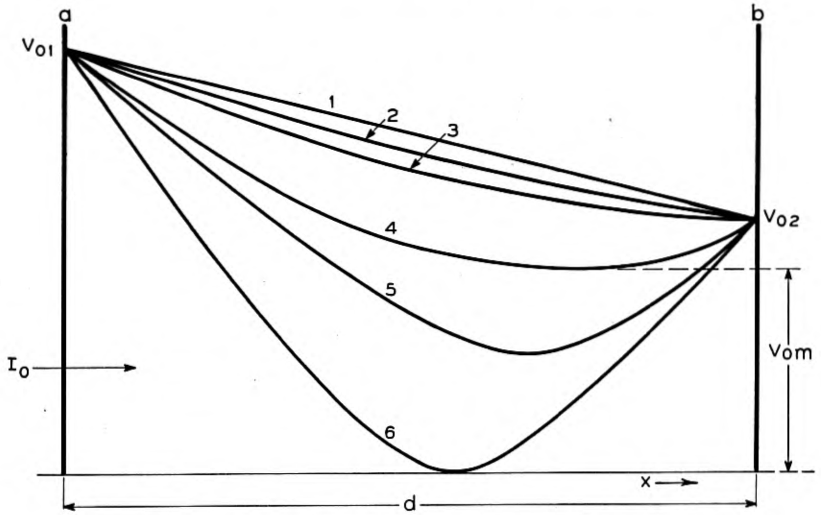


Fig. 1—Potential distributions between two planes of positive potential.

Consider two planes a and b, Fig. 1, of fixed d-c. potentials V_{01} and V_{02} ($V_{02} \leq V_{01}$) respectively, separated by a distance d. Let a unidirectional and uniform electron stream of I_0 amp./cm.[2] be injected into the space from the left at right angles to the planes. When the injected current is extremely small the potential distribution does not differ very much from the free space one, represented by curve 1 on Fig. 1. As the injected current is increased slightly, the potential curve starts to sag, curve 2, and a further small increase causes a potential minimum to develop at the electrode of lower potential, curve 3. Still more increase in injected current makes the potential

[1] Loc. cit.

minimum V_{0m} sink and move towards the electrode of higher potential, curve 4. This state of affairs, with a continuously decreasing potential minimum, continues until a critical value of injected current is reached. The potential distribution may now be represented by curve 5. The slightest further increase in injected current causes the potential minimum to sink abruptly to zero: a virtual cathode is formed. This abrupt change will be referred to as a Kipp.

With this qualitative discussion of the various potential distributions in mind a more detailed classification may be made. If both planes are assumed to be at positive potentials we may classify the different potential distributions as follows:

1. Type B
2. Type C
3. Type D

Type B corresponds to virtual cathode operation. This mode of operation will be of no interest in this paper. Type C corresponds to the case when a potential minimum at positive potential is present between the planes and type D to the case when no such potential minimum is present. For the purpose of analysis it is convenient to distinguish between two types of D solution, i.e., D_1 and D_2. This comes about because the equations for potential distribution between the planes may exhibit a minimum outside the planes. The D_1 solutions correspond to the case when no such minimum exists and the D_2 solutions to the case when such a minimum does exist.

Let us consider Type C distributions in some more detail. For this purpose attention is directed to Fig. 2. Here the ratio $\frac{V_{0m}}{V_{01}}$ is shown as a function of the injected current I_0 with the ratio $\frac{V_{02}}{V_{01}}$ as parameter. The dotted curve represents a boundary line; for currents smaller than that given by this boundary no potential minimum can exist between the planes. Consider the curve $\alpha\beta\gamma$. At the point α the potential minimum sets in and as more current is injected the potential at the minimum decreases continuously until the point β is reached. Any further increase in current causes the potential minimum to sink abruptly to zero; thus β corresponds to the Kipp point. Now it is seen that within a section of the curve $\alpha\beta$ there are two possible solutions, namely those corresponding to the section $\beta\gamma_1$ of the upper branch and those corresponding to the lower branch $\beta\gamma$. Let us designate by C_1 space charge conditions corresponding to the upper branch $\alpha\beta$ and by C_2 those corresponding to the lower branch $\beta\gamma$.

After this survey of the several possible d-c. space charge conditions we may proceed to the a-c. phenomena involved. At the present time the impedance between the two planes of Fig. 1 can be found only if the electrons move in one direction, i.e. if no virtual cathode is present between the planes, and therefore this assumption is made. It will be further assumed that electrode "a" where the injection takes place is at a-c. ground potential. This insures constant conduction current and electron speed at the plane of injection. The impedance

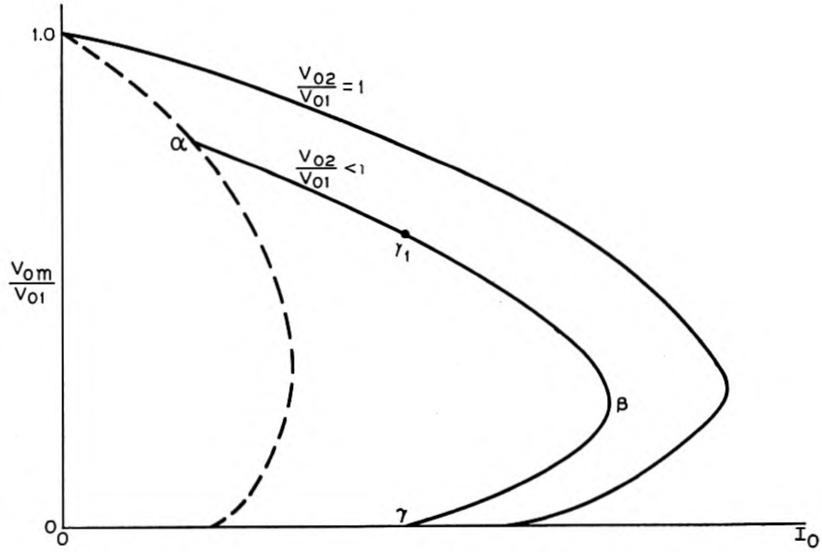


Fig. 2—Variations of the magnitude of the potential minimum as function of injected current.

may then be found by proper application of the general theory developed by Müller and Llewellyn.[2] The result is that the impedance may be represented by a series combination of a resistance r and a capacitance C having the values

$$r = J_0 \frac{T^4}{\epsilon} \cdot \frac{2 - 2\cos\theta - \theta\sin\theta}{\theta^4}$$
$$C = C_0 \cdot \frac{1}{1 - J_0 \dfrac{T^3}{d} \cdot \dfrac{2\sin\theta - \theta - \theta\cos\theta}{\theta^3}}, \tag{1}$$

[2] J. Müller, *Hochfrequenztechnik u. Electroakustik*, May 1933; F. B. Llewellyn, *Bell Sys. Tech. Jour.*, October 1935.

where

$J_0 = \frac{e}{km\epsilon} I_0$

$\theta = \omega T$

$C_0 = \frac{\epsilon}{d}$ is the cold capacitance

T is transit time, seconds
e is electronic charge, coulombs
m is electronic mass, grams

$\frac{e}{m} = 1.77 \times 10^8$ coulomb/gm.

ϵ is permittivity of vacuum $= 8.85 \times 10^{-14}$ Farads/cm.[3]
$k = 10^{-7}$.

Present interest lies mainly in the range where the transit angle θ is small. Expanding (1) and assuming θ small, we have:

$$r = \frac{J_0 T^4}{12\epsilon}$$
$$C = C_0 \frac{1}{1 - J_0 \frac{T^3}{6d}}. \tag{2}$$

These formulas are also found in Müller's paper.[2]

With regard to the resistance r it is immediately seen that it is always positive and has the same equation as for complete space charge with current I_0 and transit time T.

From the capacitance equation it is immediately evident that an increase is caused by the presence of electrons. The dielectric constant of space charge under the stipulated conditions is seen to be dependent upon the d-c. conduction current I_0. As the injected current is increased the capacitance increases initially at a fairly slow rate, but as $J_0 \frac{T^3}{6d}$ becomes comparable with unity the capacitance rises rapidly. It becomes infinite for

$$J_0 \frac{T^3}{6d} = 1 \tag{3}$$

and if the left member of (3) were to become greater than unity the capacitance would be negative. The possibility of such a condition deserves careful consideration.

[2] Loc. cit.

From the discussion of the d-c. space charge it follows that the d-c. current has an upper limit; i.e., the Kipp current with the value

$$I_{0K} = \frac{4}{9}\epsilon\sqrt{\frac{2e}{mk}}\frac{(\sqrt{V_{01}} + \sqrt{V_{02}})^3}{d^2}. \tag{4}$$

To determine the relation between the Kipp current (4) and the current (3) required for infinite capacitance, consider the d-c. equations:

$$\begin{aligned} u_b &= u_a + a_a T + \frac{eI_0}{km\epsilon}\frac{T^2}{2} \\ d &= u_a T + a_a \frac{T^2}{2} + \frac{eI_0}{km\epsilon}\frac{T^3}{6}, \end{aligned} \tag{5}$$

where u_b and u_a are electron speeds in cm./sec. at planes b and a respectively and a_a is the acceleration in cm./sec.[2] at plane a. Eliminating a_a and introducing the values of u_b and u_a in terms of V_{02} and V_{01} we find

$$T^3 - 6\sqrt{\frac{2m}{ek}}\frac{k\epsilon}{I_0}(\sqrt{V_{01}} + \sqrt{V_{02}})T + \frac{12km\epsilon}{I_0}d = 0. \tag{6}$$

When the transit time T is eliminated between (3) and (6) the result is

$$I_0 = \frac{4}{9}\epsilon\sqrt{\frac{2e}{mk}} \cdot \frac{(\sqrt{V_{01}} + \sqrt{V_{02}})^3}{d^2}. \tag{7}$$

But this is precisely the Kipp current as given by (4). Equation (3), therefore, expresses the Kipp relation between current and transit time. It has thus been found that the series capacitance at the Kipp point becomes infinite, whereas the impedance between the two planes is a pure resistance.

Consider next the possibility of $J_0 \frac{T^3}{6d}$ attaining values larger than unity when the d-c. current I_0 is limited to values smaller than the Kipp value (4). The only manner in which this could happen would be for (6) to have one root T_2 such that $T_2 > T_k$ where T_k is the transit time at Kipp. To determine this (6) is transformed by introducing I_{0K} and T_K as parameters. Thus,

$$\left(\frac{T}{T_K}\right)^3 \frac{I_0}{I_{0K}} - 3\frac{T}{T_K} + 2 = 0. \tag{8}$$

The discriminant D of (8) is

$$D = \left(\frac{I_{K0}}{I_0}\right)^2\left(1 - \frac{I_{K0}}{I_0}\right) \tag{9}$$

and as $I_0 \leq I_{K0}$ it is clear that

$$D \leq 0. \tag{10}$$

Hence, since the discriminant in general is negative (8) has three real roots. Two of these are positive and one negative. A double root occurs for $I_0 = I_{K0}$, and the value of this root is clearly $T = T_K$. For currents smaller than the Kipp current there are two positive roots T_1 and T_2 such that $T_1 < T_K$ and $T_2 > T_K$.

Thus it becomes evident that space charge conditions corresponding to the roots T_2 result in a negative capacitance. Interpreted in terms

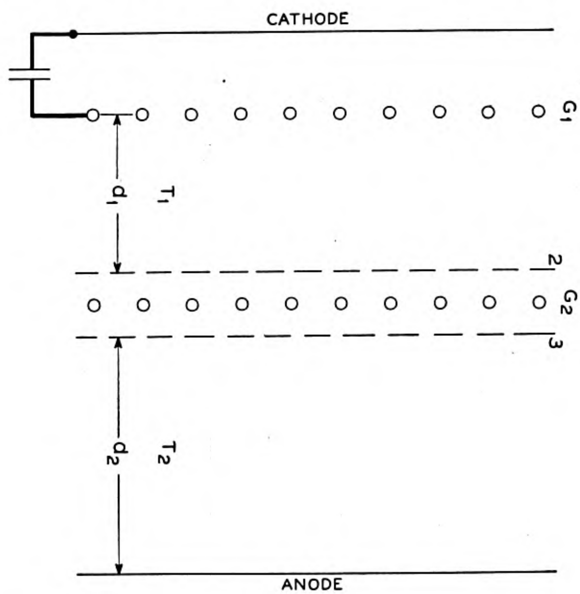


Fig. 3—Schematic of a space charge grid tube.

of Fig. 2, the roots T_1 correspond to the upper branch $\alpha\beta$ and the roots T_2 to the lower branch $\gamma\beta$. Hence, it is evident that space charge corresponding to the lower branch has a negative dielectric constant.

Consider next the system pictured schematically in Fig. 3. It differs basically from that in Fig. 1 in that a negatively polarized grid has been interposed between the planes a and b of Fig. 1. The arrangement shown in Fig. 3 may be considered to be equivalent to a space charge grid tube with the accelerating grid G_1 and control grid G_2. The grid G_1 is assumed to be by-passed to the cathode for a-c. The planes indicated by 2 and 3 are imaginary planes located on opposite sides of the control grid wires and are assumed to be sufficiently far away from the grid to insure that the potential distribution at these planes is essentially that of a grid-free space. Moreover, if the grid is fine-meshed these planes may be located quite close together so that the distance between them is negligible in comparison with the distances between G_1 and plane 2, and between plane 3 and the plate. This insures that the potentials of the two planes are practically equal. Since the grid G_2 draws only displacement current, the conduction currents at the planes 2 and 3 are equal. Since the potentials of the two planes are equal, the electron speeds are also equal. The boundary conditions in the plane of G_1 are the same as those used in deriving (1), i.e., constant conduction current and electron-speed. Then, following the method employed by Llewellyn [2] in his treatment of the negative triode, the input impedance may be found. Since the details are uninteresting the result is merely quoted. On the assumption that the plate is short-circuited to the cathode by a large condenser the input impedance between grid and cathode may be written as:

$$Z = Z_g + \frac{A_1 A_2}{\epsilon(A_1 + A_2 + G_1 B_2 + D_1 C_2)}, \tag{11}$$

where Z_g is the impedance (capacitive) between the planes 2 or 3 and G_2 and where A_1, A_2, G_1, B_2, D_1 and C_2 have the values:

$$\left.\begin{aligned}
A_1 &= \frac{1}{(i\omega)^4}\left[(i\omega)^3 d_1 + \frac{eI_0}{km\epsilon}(2 - 2e^{-i\theta_1} - i\theta_1 - i\theta_1 e^{-i\theta_1})\right] \\
A_2 &= \frac{1}{(i\omega)^4}\left[(i\omega)^3 d_2 + \frac{eI_0}{km\epsilon}(2 - 2e^{-i\theta_2} - i\theta_2 - i\theta_2 e^{-i\theta_2})\right] \\
G_1 &= \frac{eI_0}{km\epsilon u_{02}(i\omega)^2}(1 - e^{-i\theta_1} - i\theta_1 e^{-i\theta_1}) \\
B_2 &= \frac{1}{(i\omega)^3}[a_{03}(i\theta_2 e^{-i\theta_2} + e^{-i\theta_2} - 1) + u_{02} i\omega(e^{-i\theta_2} - 1) \\
D_1 &= \frac{1}{(i\omega)^2}\left[1 - e^{-i\theta_1} - \frac{a_{02}}{i\omega u_{02}}(1 - e^{-i\theta_1} - i\theta_1 e^{-i\theta_1})\right] \\
C_2 &= \frac{eI_0}{km\epsilon(i\omega)^2}(i\theta_2 e^{-i\theta_2} + e^{-i\theta_2} - 1)
\end{aligned}\right\}. \tag{12}$$

In (12):

u_{02} is the d-c. speed at planes 2 or 3.

a_{02} and a_{03} are the d-c. accelerations at the planes 2 and 3 respectively.

[2] Loc. cit.

When the transit angles θ_1 and θ_2 are very small and when the effect of the space between control grid and anode may be ignored, evaluation of (11) yields the result:

$$Z = Z_g + r_1 \frac{1 - \frac{4}{3}\dfrac{J_0 \frac{T_1^2}{2u_{02}}}{J_0 \frac{T_1^3}{6d_1}} \cdot \dfrac{1 - J_0 \frac{T_1^3}{6d_1}}{1 - J_0 \frac{T_1^2}{2u_{02}}}}{1 - J_0 \frac{T_1^2}{2u_{02}}} + \frac{1}{i\omega} \frac{1 - J_0 \frac{T_1^3}{6d_1}}{C_0\left(1 - J_0 \frac{T_1^2}{2u_{02}}\right)}, \tag{13}$$

where

$$\left.\begin{aligned} r_1 &= \frac{J_0}{\epsilon}\frac{T_1^4}{12} \\ C_0 &= \frac{\epsilon}{d_1} \end{aligned}\right\}. \tag{14}$$

C_0 is the capacitance in the absence of electrons.

The input impedance is thus seen to be a series circuit composed of a resistance and a capacitance, i.e.

$$Z = \rho + \frac{1}{i\omega C}, \tag{15}$$

where

$$\frac{1}{\rho} = \frac{1}{r_1} \cdot \frac{1 - J_0 \frac{T_1^2}{2u_{02}}}{1 - \frac{4}{3}\dfrac{J_0 \frac{T_1^2}{2u_{02}}}{J_0 \frac{T_1^3}{6d_1}} \dfrac{1 - J_0 \frac{T_1^3}{6d_1}}{J_0 \frac{T_1^2}{2u_{02}}}}, \tag{16}$$

$$\frac{1}{C} = \frac{1}{C_g} + \frac{1}{C_0} \frac{1 - J_0 \frac{T_1^3}{6d_1}}{1 - J_0 \frac{T_1^2}{2u_{02}}}. \tag{17}$$

In studying the capacitance and resistance as functions of space charge it is evident from (16) and (17) that space charge enters essentially through the functions $\frac{J_0 T_1^2}{2u_{02}}$ and $\frac{J_0 T_1^3}{6d_1}$. In what follows the capacitance C_g is assumed to be so large that the first term in (17) may be ignored. The ratio between "hot" and "cold" capacitance may then be written:

$$\frac{C}{C_0} = \frac{1 - J_0 \dfrac{T_1^2}{2u_{02}}}{1 - J_0 \dfrac{T_1^3}{6d_1}}. \tag{18}$$

Write the conductance $1/\rho$ as

$$\frac{1}{\rho} = \frac{1}{r_1} \cdot \mathrm{F} \tag{19}$$

and note that the first factor of the right member is always positive. F may be termed the relative input conductance.

Before the theoretical curves are discussed a few words about the functions $\frac{J_0 T_1^2}{2u_{02}}$ and $J_0 \frac{T_1^3}{6d_1}$ are in order. They will be expressed in terms of two parameters, i.e., φ and ξ where

$$\varphi = \frac{V_{02}}{V_{01}} \tag{20}$$

and ξ is a constant of integration which assumes different values depending upon the type of space charge present.[3] In terms of these parameters one may show that:

$$\left.\begin{aligned} J_0 \frac{T_1^3}{6d_1} &= \frac{[\sqrt{\xi^{1/2}+1} - \sqrt{(\xi\varphi)^{1/2}+1}]^3}{(\xi^{1/2}-2)\sqrt{\xi^{1/2}+1} - ((\xi\varphi)^{1/2}-2)\sqrt{(\xi\varphi)^{1/2}+1}} \\ J_0 \frac{T_1^2}{2u_{02}} &= \frac{[\sqrt{\xi^{1/2}+1} - \sqrt{(\xi\varphi)^{1/2}+1}]^2}{(\xi\varphi)^{1/2}} \end{aligned}\right\} \begin{array}{c}\text{Type} \\ D_1 \\ 0 < \xi < \infty,\end{array} \tag{21}$$

$$\left.\begin{aligned} J_0 \frac{T_1^3}{6d_1} &= \frac{[\sqrt{\xi^{1/2}-1} - \sqrt{(\xi\varphi)^{1/2}-1}]^3}{(\xi^{1/2}+2)\sqrt{\xi^{1/2}-1} - ((\xi\varphi)^{1/2}+2)\sqrt{(\xi\varphi)^{1/2}-1}} \\ J_0 \frac{T_1^2}{2u_{02}} &= \frac{[\sqrt{\xi^{1/2}-1} - \sqrt{(\xi\varphi)^{1/2}-1}]^2}{(\xi\varphi)^{1/2}} \end{aligned}\right\} \begin{array}{c}\text{Type} \\ D_2 \\ \frac{1}{\varphi} < \xi < \infty,\end{array} \tag{22}$$

[3] The parameters $1/\alpha$ and $1/\beta$ used by Fay, Samuel and Schockley are related to ξ as follows:

$$\xi = \frac{1}{\beta} \text{ Type } \mathrm{D}_1$$

$$\xi = \frac{1}{\alpha} \text{ Type } \mathrm{D}_2.$$

$$\left.\begin{aligned} J_0\frac{T_1^3}{6d_1} &= \frac{[\sqrt{\xi^{1/2}-1}+\sqrt{(\xi\varphi)^{1/2}-1}]^3}{(\xi^{1/2}+2)\sqrt{\xi^{1/2}-1} + ((\xi\varphi)^{1/2}+2)\sqrt{(\xi\varphi)^{1/2}-1}} \\ J_0\frac{T_1^2}{2u_{02}} &= \frac{[\sqrt{\xi^{1/2}-1}+\sqrt{(\xi\varphi)^{1/2}-1}]^2}{(\xi\varphi)^{1/2}} \end{aligned}\right\} \begin{array}{c}\text{Type}\\ C_1\\ \dfrac{1}{\varphi} < \xi < \left(1+\dfrac{1}{\sqrt{\varphi}}\right)^2.\end{array} \qquad (23)$$

The parameter ξ is without physical significance for the D solutions but serves merely as a convenient constant of integration. For the solutions of C type, however, ξ is of direct physical significance. For here ξ represents the ratio $\frac{V_{01}}{V_{0m}}$. In the language used in the previous qualitative discussion of space charge it is clear that $\xi = \frac{1}{\varphi}$ is the point where a potential minimum sets in and $\xi = \left(1+\frac{1}{\sqrt{\varphi}}\right)^2$ is the Kipp point. Calculations for C_2 solutions are not included.

The result of the calculations of capacitance and conductance are shown on Figs. 4 to 8. Figure 8 is an enlargement of Fig. 7 for small

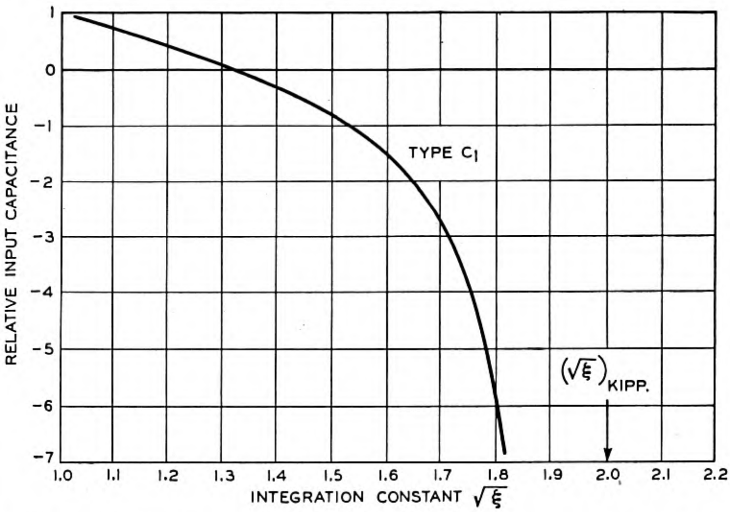


Fig. 4—Relative input capacitance of an idealized space charge grid tube for different space charge conditions (voltage ratio $\varphi = 1.0$).

values of F. Three values of the parameter φ were selected, i.e., 0.04, 0.25 and 1. The curves demonstrate regions of negative capacitance as well as of negative conductance. Note that as the parameter φ is made smaller both the capacitance and conductance pass through

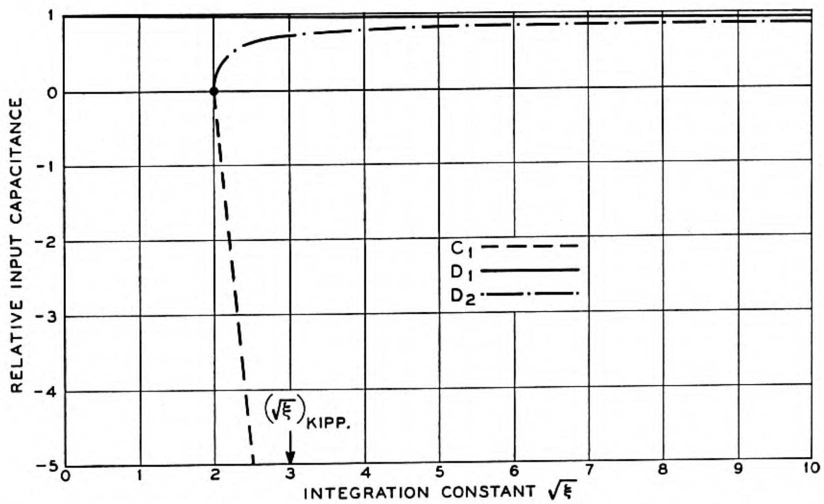


Fig 5—Relative input capacitance of an idealized space charge grid tube for different space charge conditions (voltage ratio $\varphi = 0.25$).

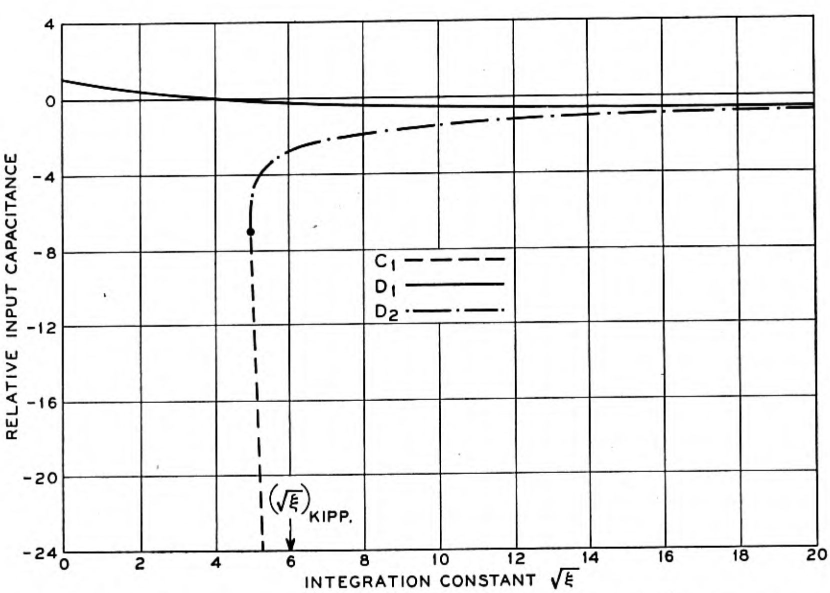


Fig. 6—Relative input capacitance of an idealized space charge grid tube for different space charge conditions (voltage ratio $\varphi = 0.04$).

zero for smaller values of ξ. Consider for instance the case for which $\varphi = 0.04$. The capacitance is here negative even in regions of small space charge and it is seen that over a wide range of the parameter ξ the capacitance is nearly constant. Both capacitance and conductance

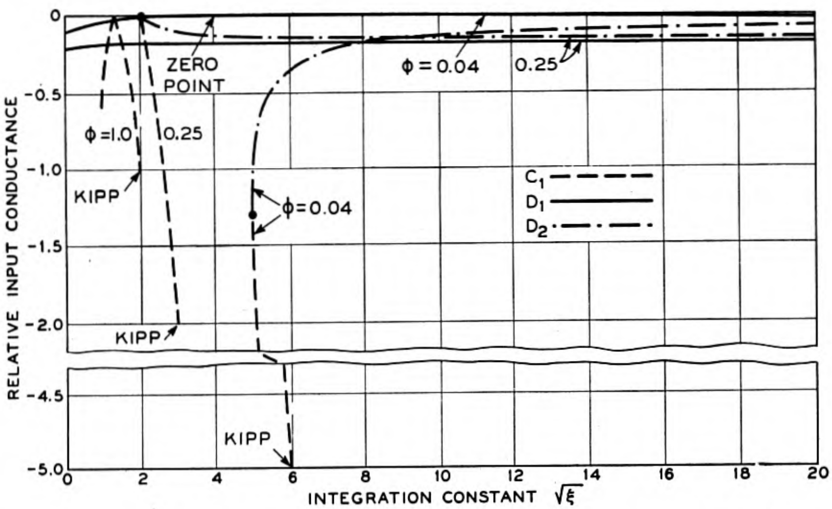


Fig. 7—Relative input conductance of an idealized space charge grid tube for different space charge conditions.

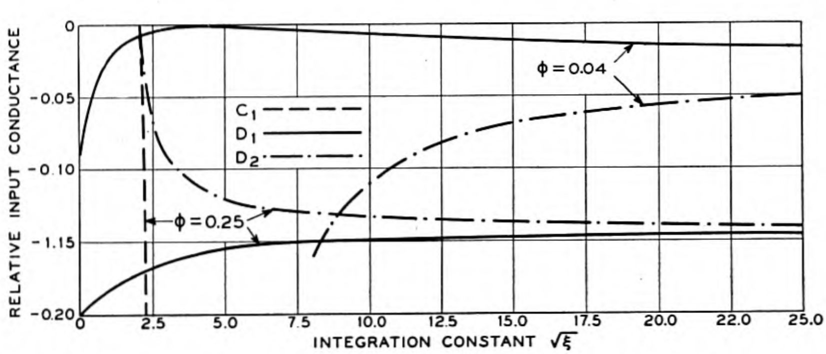


Fig. 8—Relative input conductance of an idealized space charge grid tube for different space charge conditions.

pass through zero for the same value of ξ and one sees that in case of $\varphi = 0.04$ no critical adjustment seems necessary. The conclusion is thus arrived at that the idealized space charge grid tube may be made to operate without input capacitance or loading up to moderately high frequencies. In addition, the possibility of designing an electronic negative capacitance is opened up and this is a highly desirable objective.

However, it must be kept in mind that in the discussion, effects such as velocity distributions, electron deflections and dispersion forces have been neglected. Furthermore, in an actual tube the capacitance C_g must also be considered. At the Kipp the capacitance C is equal to $-\infty$; therefore, as the Kipp point is approached a positive capacitance is expected.

The capacitance and conductance both pass through zero when

$$J_0 \frac{T_1^2}{2u_{02}} = 1, \tag{24}$$

or when

$$u_{02} = J_0 \frac{T_1^2}{2}. \tag{25}$$

But in general

$$u_{02} = u_{01} + a_{01}T_1 + J_0 \frac{T_1^2}{2}, \tag{26}$$

where u_{01} and a_{01} are the d-c. speed and acceleration in the plane of grid G_1, Fig. 3.

The relation between initial speed and acceleration for zero capacitance and conductance is, therefore:

$$u_{01} + a_{01}T_1 = 0 \tag{27}$$

and since both u_{01} and T_1 are inherently positive, the initial acceleration must be negative. For the capacitance to be negative the requirement is obviously

$$u_{01} + a_{01}T_1 < 0. \tag{28}$$

Necessary requirements for a negative capacitance are thus a finite electron speed and a retarding field at the plane of injection.

PART II

Experimental

In this section some experimental results will be discussed. The measurements all refer to the capacitance between control grid and ground of some experimental tubes. The tubes were cylindrical in structure and contained two positive grids close to the cathode followed by a negative control grid. The first positive grid has the essential function of controlling the magnitude of the current whereas the second determines the initial speed with which the electrons enter the space adjacent to the control grid.

In Fig. 9 the measured input capacitance and plate current are shown as functions of the voltage V_{g1} of the first grid. It is seen that as the plate current increases the capacitance gradually decreases, passes through zero somewhat before the plate current has reached its maximum and then becomes negative; it passes through a minimum and then gradually assumes a positive value equal to about twice the cold capacitance. This behavior of the capacitance is typical of the formation of a virtual cathode, which in the present instance appears

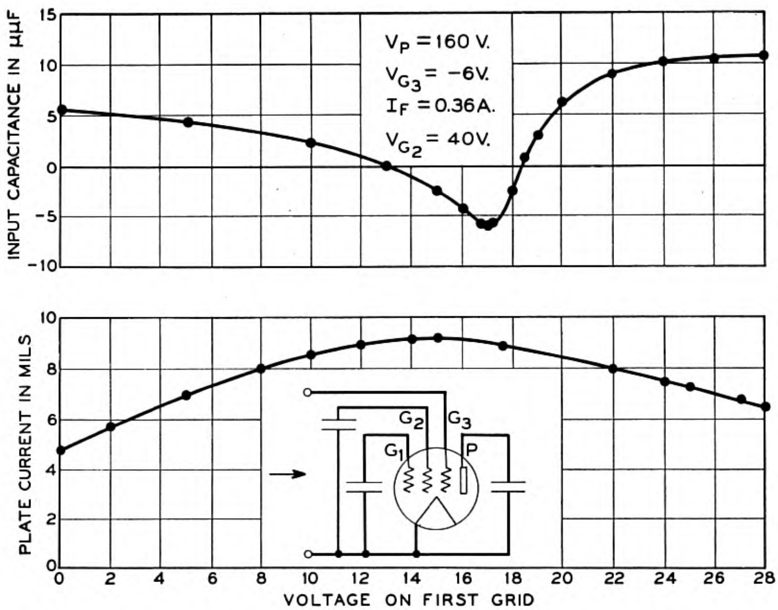


Fig. 9—Measured input capacitance of experimental vacuum tube no. 55 (grid signal = 0.18 volts r.m.s. at 50 kc.).

to be gradual. The negative capacitance is present in a small interval of the voltage V_{g1} immediately before and perhaps also during a part of the virtual cathode formation.

Figure 10 shows the measured input capacitance as a function of the control grid bias V_{g3} for several values of the voltage V_{g1}. For comparison purposes the corresponding plate currents are also shown. In the region of low plate current where a virtual cathode is present the capacitance is positive and larger than the cold capacitance. Where the plate current curve starts to bend over, that is, where the virtual cathode starts to release, the capacitance decreases. It is negative in a small domain and turns ultimately positive. The plate current curve corresponding to $V_{g1} = 30$ volts indicates that the

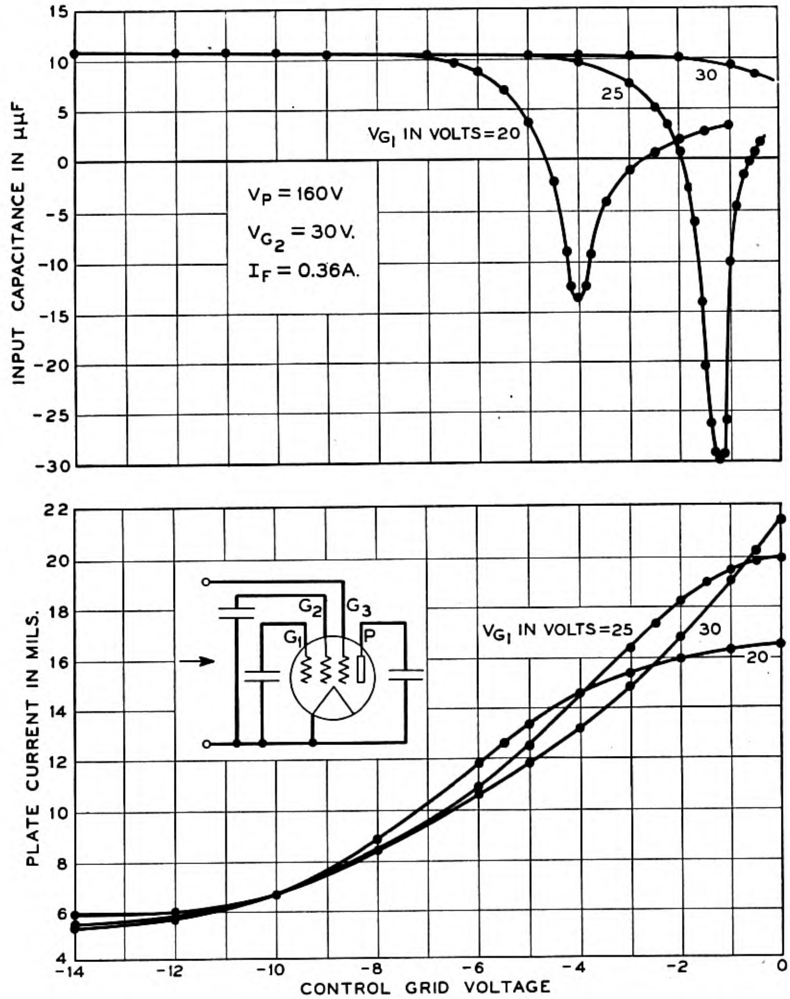


Fig. 10—Measured input capacitance of experimental vacuum tube no. 55 (grid signal = 0.18 volts r.m.s. at 50 kc.).

virtual cathode is present throughout the range of negative bias and the corresponding capacitance curve is positive everywhere.

Figure 11 refers to capacitance measurements on a tube in which a

Kipp occurred. Again the capacitance behaves essentially the same except that it suddenly jumps from a large negative value to the positive value corresponding to virtual cathode operation.

In comparing the theoretical and experimental results, it is seen that the theory gives predictions which are broadly in accord with experiments.

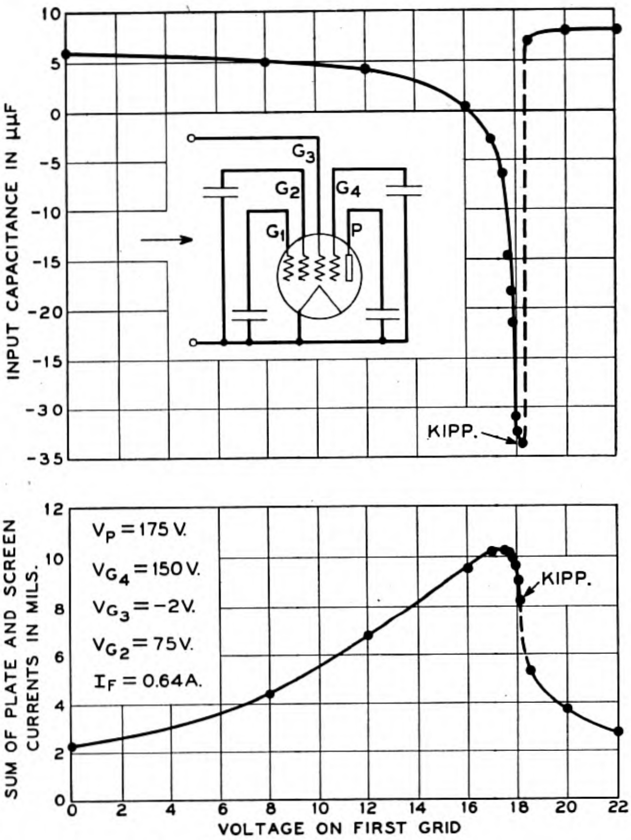


Fig. 11—Measured input capacitance of experimental vacuum tube no. 59 (grid signal = 0.05 volts r.m.s. at 50 kc.).

ACKNOWLEDGMENT

I am indebted to Mr. E. J. Buckley for assistance in the experimental work, to Miss M. Packer for the numerical calculations, to Mr. C. A. Bieling for the mechanical tube design, and last but not least to Dr. F. B. Llewellyn for stimulating discussions.

Plastic Materials in Telephone Use*

By J. R. TOWNSEND and W. J. CLARKE

ORGANIC plastics are used extensively in the manufacture of telephone apparatus and equipment. They belong to the class of materials known as insulators but are very often employed not only for their electrical properties but for their unique manufacturing and structural possibilities. Good insulating materials are very important in the telephone field although the voltage and current used are much smaller than in the power field. Progressive improvement in transmission, especially for long distance telephone service, has required that the telephone industry as a whole provide sensitive instruments and that there be a minimum loss of the electrical impulse due to leakage through insulating materials.

Rubber became at one time the most universally used insulating material in telephone apparatus. Where superior insulating properties are required, rubber has been employed not only in the soft vulcanized form as a covering for wire but as hard rubber. Its use was considerably curtailed as a molding material during the period of the world war due to the high price of rubber. This stimulated the substitution of phenol plastics which were found to produce more permanent parts. Although rubber must be classed as an organic plastic, it will not be dealt with here except in passing since it comprises a large field in its own right and quite distinct from that of the synthetic plastics. In recent years rubber has been greatly improved in life, stability, light sensitivity and resistance to cold flow so that its technical uses in the telephone plant are again increasing.

Shellac and asphalt plastics, both natural materials, were among the early important plastics employed in the telephone art. A shellac compound is still the best material for a panel system commutator where there are many long and delicate contact segments and where the principal problem is to obtain accurate location between the insulation and the brass segments together with uniform wear. The low molding temperatures and pressures for the shellac mica compound contribute to the success of the manufacture of this part. (See Fig. 1.)

Other early plastics that have found some limited uses are cellulose

* Presented before the Organic Plastics Section of the Paint and Varnish Division of the American Chemical Society, Baltimore, Md., April 3–7, 1939.

nitrate and casein. However, the cellulose nitrate plastics were only sparingly employed for telephone construction because of the serious fire hazard. Casein has long been used for key buttons and similar minor applications.

The great expansion of the use of molded plastics in the telephone plant really began with the development of organic materials which had superior manufacturing and structural characteristics over other materials. The newer plastics are of value, therefore, as much from the economies of manufacture as from their superiority over the previously used materials.

Plastics are conventionally divided into two groups: (1) the thermoplastics and (2) the thermosetting plastics. The first, considered as

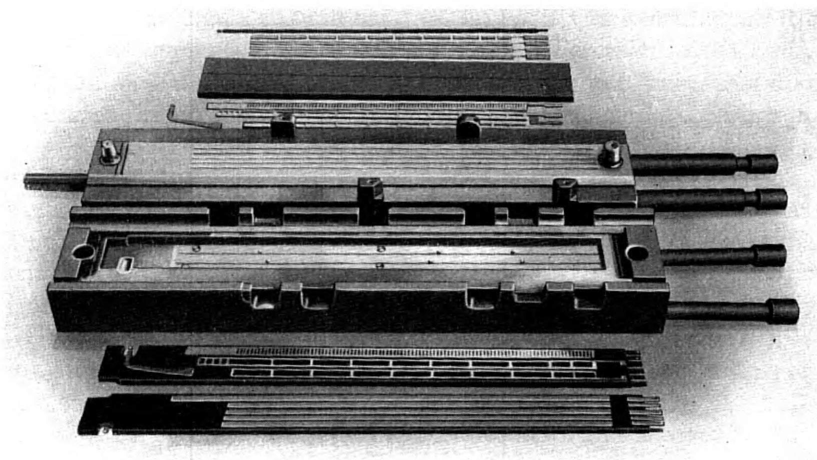


Fig. 1—Panel system commutator (flash type mold).

organic materials, are permanently soluble and fusible as well as fairly rigid at normal or working temperatures and may be deformed under heat and pressure. The second are initially thermoplastic and become insoluble and infusible after a period of time upon application of heat and pressure. These important properties are due to the chemical nature and molecular structure of the materials. All synthetic plastics are polymeric substances, that is, they are the result of a polymerization or condensation from simple organic molecules by linkage of the molecules in fairly definite ways. By a polymerization reaction is meant a reaction in which a more or less considerable number of molecules unite to form larger complexes of the same chemical composition. A condensation reaction on the other hand is one in which molecules join together to give a larger complex but during the reaction there is a separation of a small amount of water, alcohol or some other substance, so that the final chemical composition is not quite the same as at the start.

Those materials which are thermoplastic and readily soluble owe these characteristics to the fact that the molecules are linked essentially in chain-like fashion. The forces holding the chains together are of a secondary valence character, that is, the chains are free to move apart when heat is applied or a solvent is present. Vinyl, acryl and styryl resins are typical thermoplastics and incidentally each is formed from a monomeric material containing the characteristic ethylene grouping $CH_2 = C<$. The properties of synthetic materials of the thermoplastic type vary with the average chain length and distribution, and with the nature of any side groups which may be attached to the main hydrocarbon chains. The materials do not have sharp melting points as do more simple organic substances but soften gradually when they are warmed. Usually on further heating decomposition occurs to the monomeric form before any rapid flow point is reached.

The molecules of thermosetting plastics are initially in chain-like form also, although the chains are generally much shorter in length than those of the thermoplastics. On heating, the material fuses and the chains become cross-linked sufficiently to give a permanent and rigid three-dimensional structure. In the case of a phenol-formaldehyde resin the linkage between chains is directly through CH_2 groups (or according to some investigators through $-O-CH_2$ or $-CH_2-O-CH_2$-groups) and the force necessary to separate the chains is therefore high and of an order equal to that needed to break up any complex organic molecule. Infusibility and inability to go into solution are consequently the prominent characteristics of a phenolic resin in the heat-hardened form. In other materials belonging in this class the cross-linkage may be brought about through oxygen or sulphur atoms though the hardening action sometimes occurs more slowly. Certain oxygen convertible alkyd resins for example are particularly useful as organic finishes because they may be greatly hardened and in this case very much toughened by a baking process, the necessary oxygen for cross-linkage of the polymeric resin being absorbed from the surrounding air.

The properties of the more modern materials including the phenol plastics, the cellulose derivatives and the ethenoid (vinyl, acryl and styryl) type plastics, have yet to be fully evaluated but certain electrical and mechanical characteristics of these materials have already resulted in their adoption to a greater or lesser extent in the telephone plant. Phenol plastics have been employed in the molding of the regular telephone hand set, recent production being in excess of 1,000,000 units per year. Cellulose acetate is widely used in foil form. Plastics in the form of synthetic organic finishes are used for protection, decoration and insulating purposes on apparatus and equipment. For the purpose of discussion, plastics in the telephone plant may be grouped as follows:

1. Molding plastics.
2. Sheet materials (phenol fiber, acetate foil, etc.).
3. Synthetic organic finishes, adhesives and miscellaneous special items.

Objectives of Ideal Telephone Plastics

Telephone apparatus and equipment are not sold as consumption goods but the service rendered by it is sold to the subscriber. Good service means a minimum of breakdown due to replacement of malfunctioning parts, repairs and maintenance. High maintenance costs are inconsistent with the best service at the lowest cost. Uniformly high quality of materials throughout the economic life of the telephone plant is therefore essential.

The molding plastics and sheet materials account for the bulk of the plastics used in the telephone plant and the objectives of these materials are similar enough to permit them to be listed together. There are given below the general and specific properties that must be considered in such materials when they are to be used in the telephone industry. The level of quality demanded in specific properties will obviously depend on the application.

1. General requirements.
   a. Strength, hardness, toughness.
   b. Low density (to decrease mechanical inertia, aid manual use).
   c. Chemical inertness in air, or in contact with other materials.
   d. Resistance to humidity (minimum of swelling and shrinkage with variations of moisture content of the air).
   e. Ability to withstand temperature, heat and cold without too great impairment of strength and shape.
   f. Ability to reproduce die surface accurately and give good appearance to finished part.
   g. Light stability.
   h. Relative non-inflammability.
   i. No odor, no harm to the skin.
   j. Resistance to insect attack.

2. Specific mechanical properties.
   a. Transverse strength.
   b. Impact strength.
   c. Cold flow.
   d. Shrinkage.
   e. Wear resistance.
   f. Machinability.
3. Specific electrical properties.
   a. Insulation resistance
      1. as affected by humidity.
      2. as affected by light.
   b. Dielectric constant.
   c. Power factor.
   d. Dielectric strength.
4. Moldability.
   a. Free flow at moderate temperature and pressure.
   b. Favorable setting characteristics.
   c. Short molding cycle.
   d. No tendency to stick to die.
   e. No abrasion of die surface.
   f. Minimum shrinkage in mold.
   g. Low bulk factor.
5. Economic considerations.
   a. Low density materials preferred.
   b. Cost.
   c. Die life.
   d. Utilization of scrap.
   e. Molding cycle time.
   f. Trimming and finishing characteristics.
   g. Refinishing or maintenance.

The objectives of an ideal plastic in the telephone industry depend upon the use to which the material will be put in the telephone plant. The material may be a structural member, an insulator, or both, and may be in the hands of the public or in a telephone exchange. All of the above requirements need not be met but excellence in a majority of these properties is generally desirable.

The Molding of Telephone Parts

Molding involves consideration of (1) the molding compound (2) the die (3) the press (4) the heating and cooling system (5) method of ejecting part from mold (6) finning and trimming methods. Regardless of the type of plastic, these operations are necessary.

As to the molding compounds, the Bell System obtains from the suppliers whatever materials are needed and this is generally true for the industry. These range from the plain wood flour-filled phenolics to the various thermoplastics depending on the application. Exact compositions are seldom specified in order that the manufacturer be given all possible opportunity to exercise his ingenuity to produce satisfactory quality material.

The die or mold is such an important item in the molding of a material that several points should be emphasized about it. The dies are always expensive. Every part is an individual design problem, involving flow of material in the cavity, use of inserts, opening and closing, the clamping of die parts under high pressures, and alignment. Everything possible must be done to reduce the complication of the die; eliminate inserts if possible, provide generous fillets, ample taper for removal of parts, and facilitate flow of the compound. The Bell System has found it advantageous to make most of its own dies. The conventional boring, milling and hobbing processes are used. Very little success has been had with other methods, such as casting with hard alloys.

In spite of the expensive and time-consuming effort that must be encountered in designing and building a molding die, the finished die when properly used represents one of the most indestructible tools of modern manufacture. The parts are finished, require little or no surface treatment, the dimensions are accurate and sub-assembly operations may already be completed as the part emerges from the die.

The dies used are of the three general types: the open or flash die, the closed or positive die, and the injection die. The open type consists of two parts which come together at a cutting edge or ridge that surrounds both halves of the die. Since this ridge or cutting edge must withstand the full pressure of the press it is usually about one-eighth of an inch wide. Flash dies are all relatively simple, readily loaded, and the charge need not be measured accurately, any excess being forced out as the two halves close until the cutting edges come into contact. Such molds may be used for any molding compound not requiring high pressure and which does not have high die shrinkage. Shellac-mica commutators (Fig. 1) are manufactured by means of a flash die.

The positive type die consists of a plunger and a cavity shaped to produce the finished part. The plunger may aid in shaping the part. Only enough material is placed in the die to make the part. The material may be weighed in separate charges or preformed by a separate operation. The cavity is loaded and the plunger forced down, forming the part. Multiple cavity molds cannot be loaded exactly alike and hence some provision must be made for the escape of excess material, forming a flash that must be subsequently removed. Such dies must be carefully designed so that the fin is located for easy trimming and to provide the best appearance. These multicavity molds are frequently called semi-positive molds to differentiate them from a truly positive mold where there is little flash. A typical telephone part made in a semi-positive mold is the handset handle shown in Fig. 2.

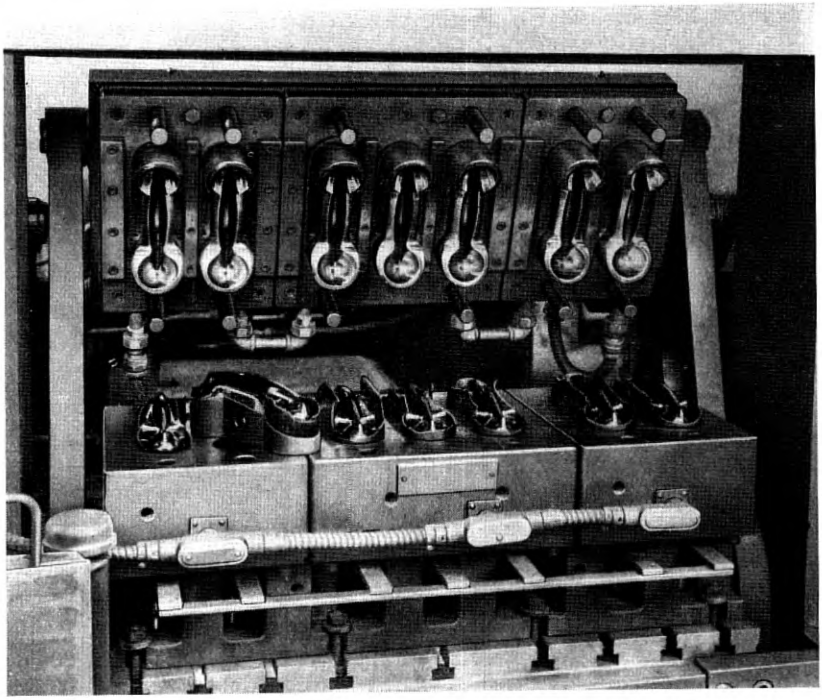


Fig. 2—Mold for handset handle.

The recently developed process of injection molding which consists in forcing plastic material through a nozzle into a completely closed die from an external compression chamber is also being used and promises to alter many of the present operations. Since the opening and the closing of the die are not related to the application of pressure, greater freedom of design is possible. Furthermore, since full pressure is not applied by this method until the die cavity is completely filled with material, the material already in the die tends to support inserts and hold them in their true position. Hence more delicate inserts are possible by this method of manufacture. Since the material is enclosed in an auxiliary pressure chamber and is not exposed to the atmosphere, greater freedom from room dust is possible, rendering this method ideal for colored plastics. A terminal block used to terminate the subscribers telephone cord is made from thermoplastic molded cellulose acetate compounds as shown in Fig. 3.

Finishing and trimming methods are largely determined by the design and the class of service required of a molded part. In the case of the injected cellulose acetate terminal block mentioned above the gates are trimmed off by a simple trimming punch and the scrap is

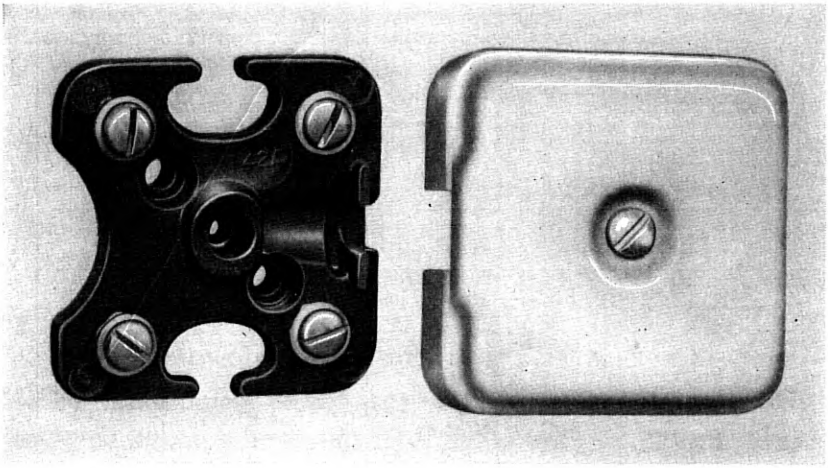


Fig. 3—Terminal block.

reused. This part is covered in service and the appearance of the block is therefore not a major factor.

The telephone handset, since it is in the hands of the public, must be carefully finned for two reasons: (1) to avoid surface roughness and (2) to provide good appearance. The handset handle was originally ground along the fin left by the semi-positive mold and then buffed. The operation was not only expensive but tended to grind off a large portion of the surface of the handle. This removed the resin-rich surface and tended to expose the filler of the phenol plastic molding compound, thus reducing the appearance life of the handle. The more recent product of the Bell System is being grooved along the die parting line. This removes the fin, a minimum of the resin-rich surface and does not detract from the appearance of the handset. Automatic grooving machines were developed for this purpose. Figure 4 shows a grooved handset handle.

It has been found necessary to pay close attention to the design in order that die parting lines, ejector pin marks, gate marks and the like will appear at points where they may be readily eliminated by simple trimming and grooving operations, or where they may be left without objection to appearance or function of the part.

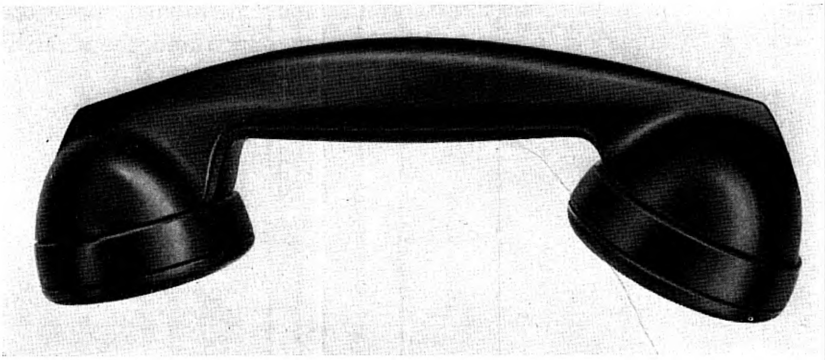


Fig. 4—Grooved handset handle.

General Test Methods and Requirements

The most satisfactory test is one that can be applied to the finished part to measure the ability of that part to perform its function satisfactorily in service. This ideal is seldom realized, not only because of the difficulty of defining the service requirements but of finding tests that are wholly representative of service conditions. It is customary, therefore, to apply a series of tests whose sum total will approach the ideal as nearly as practicable. Molded organic plastic parts are different from parts made from most other materials in that the molding process may modify them and render them quite different from the raw material. In the case of thermosetting compounds this is particularly true.

Tests are in the main applied, therefore, to a molded part of representative specimen of the fabricated material. In the telephone plant the items that are of most importance are strength, both transverse and impact, permanence of form, appearance, effect of moisture and drying on swelling and shrinkage, insulation resistance, electrical breakdown potential, and reaction on adjacent materials. Methods of making all of these tests have been worked out and are supplemented by apparatus tests made on the manufactured product.

There is no known test that will measure completely the quality of a phenol plastic part and it is necessary to resort to the expedient of testing standard bars molded under specified conditions. Five bars are molded in a positive type die as shown in Fig. 5. The step arrange-

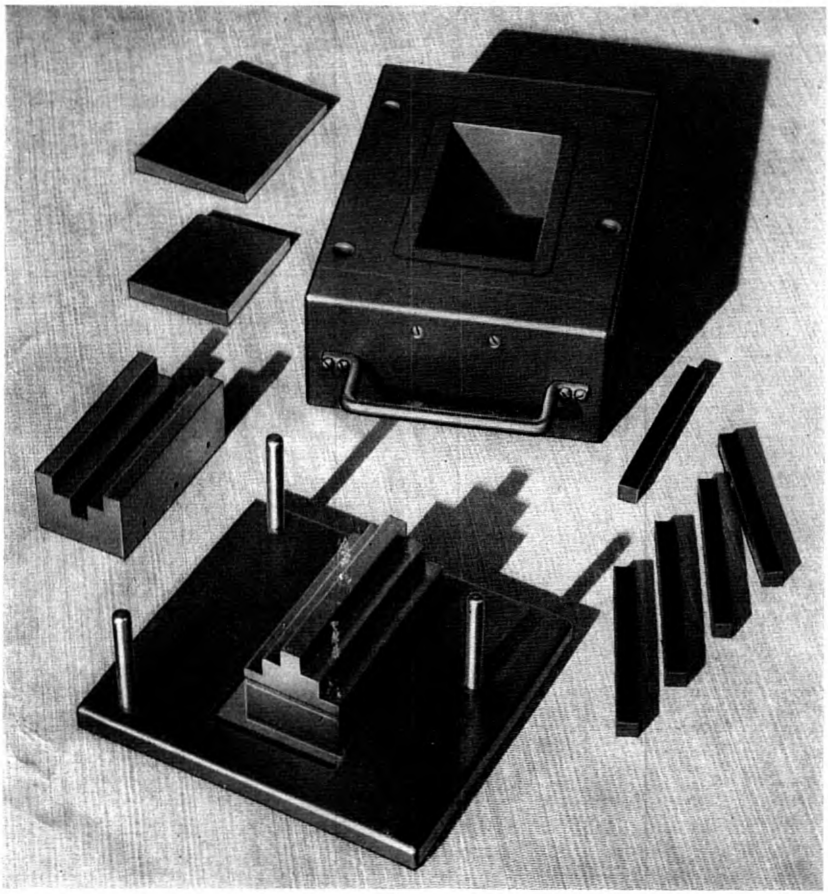


Fig. 5—Specimen mold and bars.

ment provides for flow within the die similar in many respects to the molding of actual parts. The test bars provide transverse or flexural strength, impact insulation and cold flow test specimens. The methods used, whenever possible, are those of the American Society for Testing Materials.

The cold flow test is specially designed to note the distortion of materials which in service are under pressure, such as spring pileups, inserts and apparatus in the form of banks and terminal blocks. The specimen which is ½ × ½ × ½″ in size is first conditioned at 150° F.

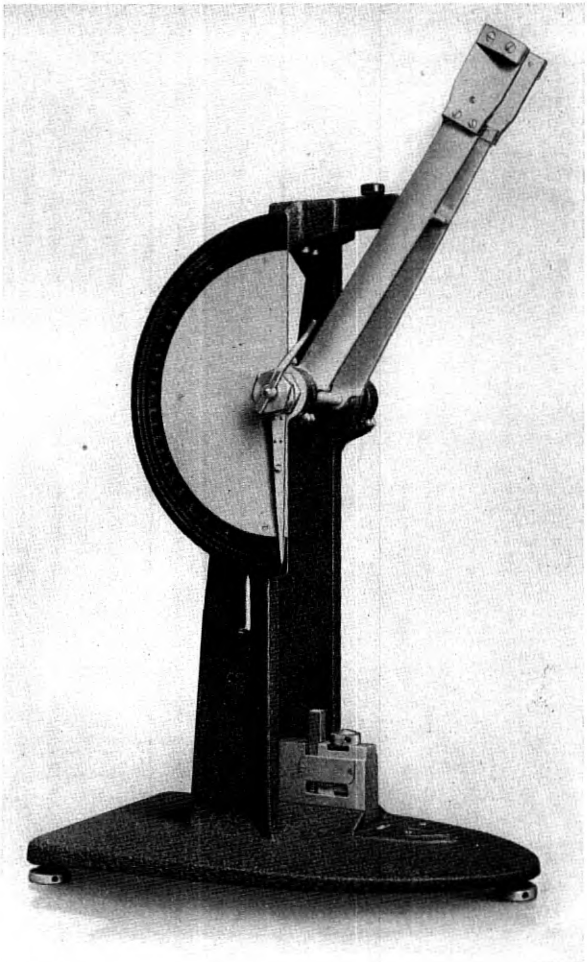


Fig. 6—Impact machine.

for 4 hours and then to 90 per cent R.H. and 85° F. for 68 hours to permit absorption of moisture. It is then held under 4000 lbs. per square inch for 24 hours at 120° F. ± 1° and the percentage change in thickness determined.

Typical Applications of Thermosetting Plastics

Phenol Plastics

A typical phenol plastic telephone part is the handle for the handset. This part is molded in a multiple semi-positive die of the kind described above.

In addition to the laboratory tests on the raw material, samples of the molded handles must withstand a dropping test. After being conditioned, handles representative of a given lot are equipped with transmitters and receivers and dropped down a nearly vertical chute to strike on a steel block. The test is made by dropping first at 36 inches and then increasing the drop in increments of 2″ until the handle breaks.[1] Normal product handles will withstand a drop of 55 inches on a steel block without failure.

The electrical properties of phenol plastic compounds are adequate for most uses in the telephone plant. Two grades are recognized, however, the mechanical and the electrical. Fully 90 per cent of the uses involve the mechanical grades. For certain high-frequency insulation purposes special mica-filled phenol plastics are used in place of the regular wood and cotton filled varieties.

One of the outstanding disadvantages of a phenol plastic is the ease with which it carbonizes on exposure to electrical arcing. For this reason phenol plastic compounds have only a limited use for commutators and similar applications. However, in addition to handsets they have proved of value for mouthpieces, receiver cases, subset housings, non-magnetic coil forms, coil cases, jack mounting blanks and terminal blocks.

Phenol Fiber

Phenol fiber for telephone apparatus is made of alpha cellulose paper, Kraft paper and rag paper by the usual impregnation with a suitable phenol resin varnish and lamination of a number of sheets under heat and pressure. The most important requirement for the paper is that it shall be pure, clean and free from electrolytes. The paper is carefully tested for chlorides, conductivity of water extract and for alcohol soluble materials. Several grades of phenol fiber are necessary to meet the requirements, some of which are largely mechanical and others electrical.

The principal tests for phenol fiber are cold flow and shrinkage, insulation resistance, corrosion tendency, arc resistance, transverse strength and impact. Arc resistance applies to the case where wiping

[1] "The Impact Testing of Plastics," Robert Burns and Walter W. Werring, *Proc. A.S.J.M.*, 19, Vol. 38, 1938.

contacts cause an arc to flash across the surface of the phenol fiber. This is a condition peculiar to telephone apparatus and a special test has been designed which simulates service conditions. An insulator cam (see Figs. 7 and 8) is prepared as the test specimen. The cams are rotated at a speed of 10.5 to 11.5 revolutions per minute. Two metal cams are attached concentrically with each surface of the phenol fiber cam. Attached to a brush that wipes over the cam tensioned to

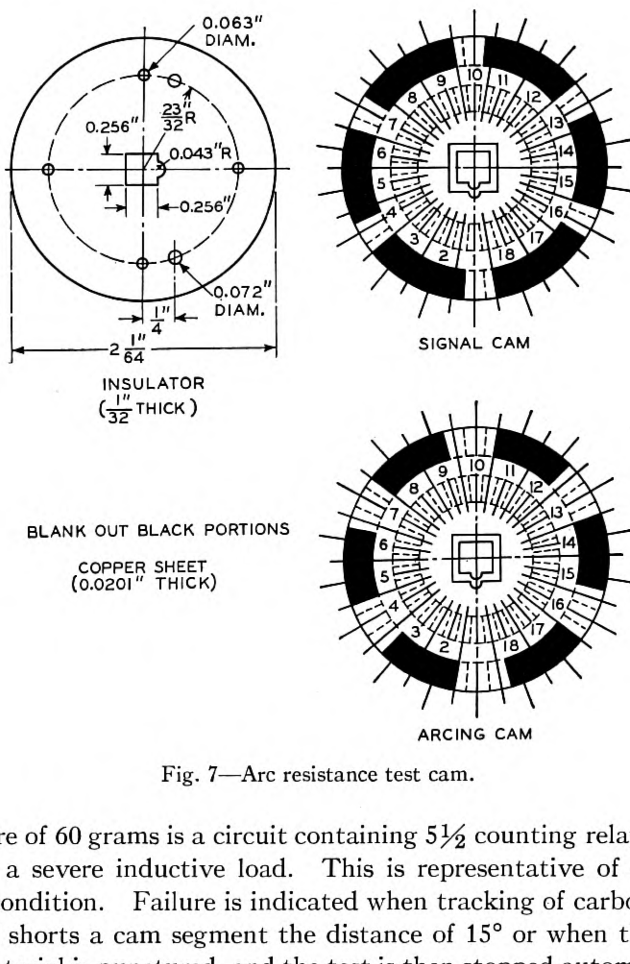


Fig. 7—Arc resistance test cam.

a pressure of 60 grams is a circuit containing 5½ counting relays which supplies a severe inductive load. This is representative of a severe service condition. Failure is indicated when tracking of carbonaceous material shorts a cam segment the distance of 15° or when the 1/32″ thick material is punctured, and the test is then stopped automatically. A good grade of fiber will resist over 1,800 revolutions whereas a poor grade will fail in 4 to 100 revolutions.

Phenol Fabric

Phenol fabric is similar to phenol fiber except that it is made with fabric instead of paper. It is generally used for its mechanical strength since its electrical properties are inferior to fiber. Phenol fabric is used in tools where high strength and resistance to impact and bending

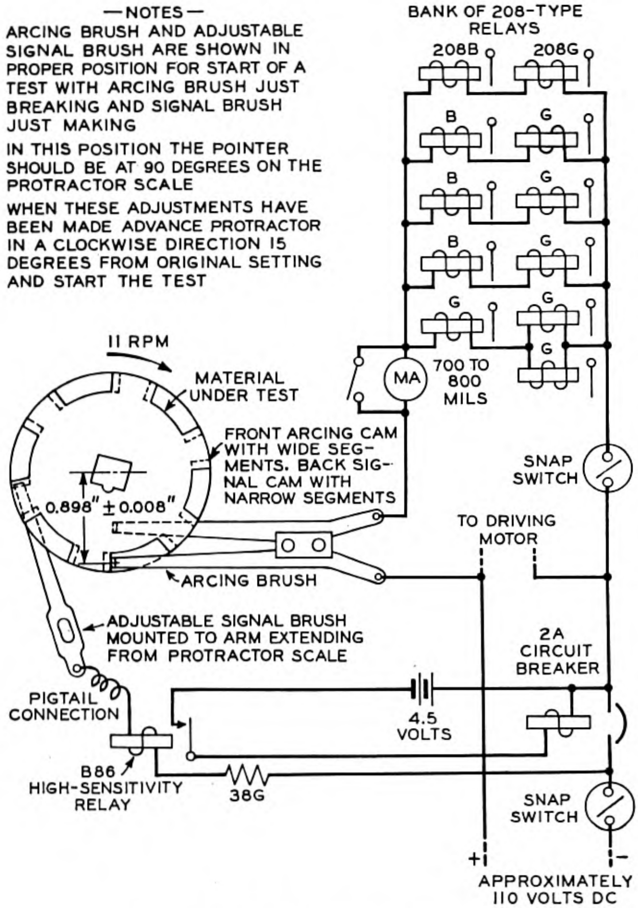


Fig. 8—Circuit for arc resistance test.

are necessary, in terminal plates, cable terminals, gears and in general where phenol fiber does not have suitable structural properties.

Urea-Formaldehyde Plastics

Condensation products made from urea and formaldehyde have attractive possibilities as thermosetting plastics. Relatively lightfast colored parts with porcelain-like surface luster are possible with these materials. The molding cycle is slightly shorter than for phenol plastics and the material at first is somewhat more fluid. Tight molds are therefore necessary in order to get sufficient pressure.

The principal difficulty with urea-formaldehyde plastics has been that on exposure to heating and cooling or humidification and drying cycles, there is a tendency toward cracking, particularly at changes in section and around inserts. Molded into uniform thin sections without inserts they are reasonably satisfactory plastics.

At present there are practically no urea-formaldehyde plastics employed in telephone apparatus because the wide continental climatic conditions and exacting requirements will not permit their use. Recently there have been improvements made from a stability standpoint and it is believed future application may be found for these plastics, particularly in view of their color permanence.

Applications of Thermoplastic Materials

Cellulose Acetate

The principal use of cellulose acetate is for interleaving in coils. Various relay coils are made of layers of cellulose acetate over which a layer of enamel coated copper wire is wound. Layer upon layer of wire and acetate sheet form a coil. These are then assembled on a core and spoolheads attached. One of the phenol fiber spoolheads has a surface coating of cellulose acetate and the winding is pressed against this spoolhead and dipped in acetone. This dissolves or softens the exposed edges of acetate and the whole coil is firmly secured to the spoolhead.

Cellulose acetate is used for this purpose since it is practically inert as regards corrosion of the fine copper wire in contact with it and in this respect it is superior to any known material. It is permanent, reasonably fireproof and has high insulation resistance. Two grades of cellulose acetate sheet are used in telephone practice. These are the window grade used as a window or covering over designations and an electrical grade for coil use. The principal tests for the electrical grade are insulation resistance, shrinkage, and resistance to burning.

The principal use of molding grade of cellulose acetate is for the terminal block (Fig. 3) mentioned above. Here the application is mainly structural since it has more than adequate electrical insulation. Another application for cellulose acetate is a test strip where a surface layer of acetate over phenol plastic avoids carbonization of the latter.

Acrylate Resins

The clear water-white plastics derived from the polymerization of the esters of acrylic and methacrylic acid are at present being used only for windows, viewing lenses on designation strips and other optical uses in telephone apparatus. This is a new plastic and applications will no doubt develop in time, taking advantage not only of color but of the mechanical and electrical properties and insensitivity to moisture.

Vinyl Resins

Vinyl acetate and vinyl chloride polymers and co-polymer products form interesting thermoplastic resins. Their use in communication work has been limited so far to phonograph records, where the resistance of the co-polymer to warping due to humidity and the superior wear resistance of the plastic have been the important factors. The advantage of non-inflammability imparted by chlorine is more than offset by the acidic nature of the fumes given off from vinyl chloride polymers when heated or burned and this has discouraged use of these materials in the telephone plant.

Polystyrene

There have been no important commercial applications for polystyrene as yet in Bell System telephone communication although much experimentation is being carried on with polystyrene plastic. The low electrical losses of this material make it of special interest in high-frequency work but its mechanical properties have not been satisfactory. In Germany and Italy it is reported that polystyrene has been employed as an insulating plastic in various cable structures for experimental and commercial use. However, in this country the spacing insulators for the coaxial cable from New York to Philadelphia have been made of a special grade of hard rubber which has proved to be a tougher material for the purpose.[2]

Synthetic Coatings

An important application of synthetic organic materials in the telephone plant is in the finishes that are put on Bell System apparatus. There are three major reasons why such finishes are needed—(1) for the improvement in appearance of certain fabricated parts, especially the exposed portions of subscriber station apparatus, both in private homes and public places, (2) for the mechanical and chemical protec-

[2] "Systems for Wide-Band Transmission over Coaxial Lines," L. Espenschied and M. E. Strieby, *Bell Sys. Tech. Jour.*, October 1934; and *Elec. Engg.*, Vol. 56, 1937.

tion of the underlying structural material which is usually a metal, and not infrequently, (3) for electrical insulation purposes. Other minor reasons for finishes exist on special apparatus. Many parts are fashioned from such metals as steel, brass, aluminum and zinc alloys. After such practices as punching and die casting the surfaces of these parts are left in an unsightly condition, and furthermore unless protected, they may soon begin to corrode. In certain cases electroplated finishes may be employed to advantage, but organic finishes, because of their low cost and ease of application, find wide use. A good organic finish for telephone apparatus must not only have a lasting decorative value but must also protect the parts against the great variety of conditions to which the apparatus is exposed.

For example, the common black finish which is applied to various parts of subscriber station apparatus, such as the zinc alloy handset mounting, coin collector boxes and metal bell boxes must be sufficiently tough and adherent to withstand perspiration, impact and severe abrasion. Rigorous tests have been applied to find the most durable finishes for such parts. They must maintain their appearance so as to harmonize with the smooth molded black phenol plastic parts. Advantage has been taken of the recent improvements in synthetic resin finishes and a modified alkyd resin vehicle has been employed in the present black enamel. A thorough baking is given to the enamel which results in a more durable finish for telephone apparatus than the former black japan.

There are a number of applications for synthetic finishes where corrosion protection is important from the standpoint of the proper functioning of the apparatus. The aluminum diaphragms in marine and aviation loud speakers and in sound power instruments are protected by a baked finish containing a heat-hardening phenolic resin vehicle into which is incorporated a chemically inhibitive pigment.

The familiar olive-green finish applied to the metal lining of telephone booths is also a synthetic finish. This coating is often subjected to unusual service conditions which only a modern type of finish can withstand. Advantage is also taken of the high initial reflectivity and the retention of light reflection of certain alkyd resins and these are used in the white booth head-lining enamel. Synthetic finishes are generally specified for the finishing of Bell System trucks, etc.

Other applications of finishes include lacquers and wrinkled enamels. A recent interesting development has been the use of ethyl cellulose dipping lacquers to form a continuous, fairly thick, envelope around small telephone parts such as resistances, condensers and the like.

This frequently eliminates a potting operation and provides an excellent mechanical protection for the parts.

The employment of organic coatings for insulation purposes is another important application deserving mention. In general the conditions in the telephone plant are not such as to demand resistance to very high voltages. Millions of feet of copper wire receive a baked clear enamel coating applied in multiple thin coats to assure maximum flexibility and uniformity. Carefully chosen pigmented alkyd baking enamels are used in exchanges as insulating finishes for various hooks, bars and other small parts of the metal framework upon which the exchange wiring is tightly and compactly fastened and from which electrical insulation is needed.

An appreciable amount of clear cellulose acetate lacquer is at present used on switchboard wire. This is applied as a thin coating of a specially plasticized lacquer over a layer of textile insulation, the latter being colored in various ways for ready identification. The requirements of a good lacquer coating material for switchboard wire are chiefly (1) low cost, (2) reasonably good insulation, even under prolonged high humidity conditions such as occur during the summer months in many parts of the country and (3) good transparency so as not to alter the identification colors on the textile serving. Smoothness, flexibility, inflammability and corrosion hazard are other important factors that receive consideration.

A synthetic plastic which has recently found a small but important place in the telephone plant is polybutene. When coated on fabric this plastic has given an excellent membrane material for a new type of handset transmitter. It is dust and moisture-proof, light in weight, flexible and alkali resistant, not impairing in any way the acoustical properties of the instrument.

Synthetic Resins as Adhesives

A growing use for synthetic resins in the telephone plant is in the form of adhesives. The amount of material consumed in this way is not large but the applications are frequently important from the standpoint of the functioning of the apparatus as well as from the economies involved. The older kinds of adhesives such as casein and animal glue are still employed for joining together various large parts (especially wood, as in cabinet work, etc.) but they are brittle and generally unsatisfactory in the assembly of small light parts (metal, phenol fiber, ceramic, etc.) such as go into special communication apparatus.

The trend in the design of most apparatus has been toward smaller lighter parts and at the same time toward more rapid assembly. Synthetic resin adhesives are aiding this trend by avoiding in various places the dependence upon bolts, screws and similar mechanical locking devices. Proprietary resin-cellulosic lacquer adhesives and vinyl and acrylate polymers are proving of value because they give strong tough joints that are affected but little by moisture and are not apt to give trouble from corrosion or growth of mildew. The use of these materials for assembling parts in a thermoplastic manner looks particularly encouraging. When the surfaces which are to be joined are carefully cleaned, then primed with an air-dried coat of a suitable thermoplastic resinous adhesive and finally molded together under heat and pressure, tensile strengths of several tons per square inch are possible between the joined parts. Synthetic resin cements and adhesives are employed in the construction of the handset transmitter, moving coil microphones, loud speakers, switchboard lamps, vacuum tubes and the wood veneer of telephone booths.

Conclusion

Many important applications of plastics have been made in the telephone field. These have sprung from the economies of design, methods of fabrication, as well as from the excellent serviceability of the molded plastic products. It might be well to emphasize again the chief limitations of present day plastics which have prevented wider use. When exposed to outdoor conditions which involve the effect of temperature and sunlight, many plastics, particularly the newer thermoplastic materials, are revealed to be insufficiently permanent for telephone use with respect to the physical and chemical characteristics associated with color, distortion at elevated temperatures, surface deterioration due to action of sunlight and brittleness at moderately low temperatures.

Modern trends in stylized designs make it necessary to take advantage of the molding art in order to achieve good ornamentation. This will probably result in the use of plastics on surfaces exposed to light. The effect of light is not confined to direct sunlight for it has been found that daylight filtered by ordinary window glass for long periods will cause reduction in surface electrical resistance of plastics. In fact, the effect of light seems to be over a rather broad range of the spectrum, becoming intensified as the ultra-violet range is approached.

A few of the new thermoplastics are quite inflammable, a characteristic that will be a serious handicap to their extended use in exchanges and other locations where a fire might disrupt the service of a whole community.

There are still certain weaknesses of organic finishes with respect to impact abrasion, perspiration and moisture penetration, although there have been great advances made in this field in recent years.

The Dielectric Properties of Insulating Materials, III
Alternating and Direct Current Conductivity

By E. J. MURPHY and S. O. MORGAN

This paper deals with the variation of a-c conductivity with frequency and with that of apparent d-c conductivity with charging time for dielectrics exhibiting anomalous dispersion (i.e., having dielectric constants which decrease with increasing frequency). The a-c conductivity of a dielectric exhibiting simple anomalous dispersion approaches a constant limiting value γ_∞ as the frequency increases. The discussion shows that γ_∞ possesses properties similar to those of the conductivity due to free ions, although in most cases it depends upon the motions of polar molecules or bound ions. It is also shown that the apparent conductivity for constant (d-c) potential approaches an initial value as the charging time is diminished. This initial conductivity γ_0 is demonstrated to be equal to the limiting value of the a-c conductivity attained at high frequencies (γ_∞), a relationship which simplifies the description of the behavior of dielectrics exhibiting simple anomalous dispersion. Dielectrics possessing the property of anomalous dispersion then have *two* conductivities: one is due to local motions of polar molecules or bound ions; the other is due to the migration of free ions to the electrodes.

Both γ_0 and γ_∞ refer to methods of measurement. It is to be noted that in many non-homogeneous dielectrics, especially those in which one part is of much higher resistivity than the remainder, both γ_0 and γ_∞ may be a measure of a free ion conductivity. As the equality of γ_0 and γ_∞ is independent of the nature of the polarization responsible for them, experimental agreement between a-c and d-c measurements cannot be used to distinguish whether the dielectric loss in a material is due to polar molecules, to bound ions, or to free ions present in a non-homogeneous dielectric. However, in homogeneous dielectrics γ_0 (or γ_∞) is a conductivity due to polar molecules or bound ions.

Introduction

THE preceding paper [1] dealt with the dielectric constant, showing mainly how it varies with the frequency of the applied alternating voltage for those dielectrics which behave in the simplest manner, and indicating the general character of the structural features responsible for this behavior. The discussion is extended here to the conductivity,

[1] Murphy and Morgan, *B. S. T. J.*, *17*, 640 (1938).

which is not less important than the dielectric constant as a property of an insulating material. Though general aspects of the conductivity will be described for the sake of completeness, we wish mainly to show that materials which possess the property of anomalous dispersion may be considered to have *two* quite definite conductivities: one of these is the ordinary d-c conductivity due to free ions or electrons; the other is a special value of the a-c conductivity which will be discussed in this paper. We believe that the recognition of the existence in many materials of two conductivities instead of one is of considerable advantage, particularly in interpreting the behavior encountered in direct-current conductivity measurements on insulating materials, a subject upon which there has existed a considerable divergence of opinion.

The measurement of the direct-current conductivity of an insulating material is usually complicated by the fact that the current which flows when a constant potential is applied does not remain constant but decreases with time. The meaning of this variation of the current is open to more than one interpretation. Some investigators consider that its final value, approached asymptotically, and perhaps not closely approximated until a constant potential has been applied for an hour or more, is the proper basis for the calculation of the true conductivity of the material. Other investigators, notably Joffé, consider that the current/time curve should be extrapolated toward the instant of applying the voltage in order to obtain the proper value of the current to use in calculating the true conductivity. On this account the terms *initial conductivity*, *final conductivity* and *true conductivity* frequently appear in papers on the conductivity of insulating materials. While it has been usual to take either the initial or the final conductivity as the true conductivity, rejecting the other, it is shown here that with certain exceptions both conductivities are true conductivities in the sense that they are independent properties of the material having a different, though related, physical significance.

The relationships which will be brought out here depend in an essential way on the nature of the variation of a-c conductivity with frequency for materials which possess the property of anomalous dispersion. The a-c conductivity of a dielectric exhibiting simple anomalous dispersion increases as the frequency increases until the frequency is high as compared with the reciprocal of the relaxation-time; it then approaches asymptotically a constant limiting value. It is shown here that this limiting value of the conductivity, which will be referred to as the *infinite-frequency conductivity*, is a true conductivity of the material, analogous to the ordinary d-c conductivity, and that it is equal to the initial conductivity obtained by extrapolating the apparent d-c conductivity towards the instant of applying the measuring voltage. We believe that this relationship considerably simplifies the description of the meaning of certain types of measurements upon dielectrics.

In spite of the fact that several terms are already used to distinguish different conductivities, there remains some ambiguity in the meaning of these terms. For example, the physical meaning of the term d-c conductivity when applied to a dielectric is vague. Moreover, it will be evident in the later discussion that the initial conductivity will depend upon free ions for some materials and upon polar molecules for other materials. To avoid this confusion we have found it convenient to use two terms which refer to the nature of the conduction processes rather than to the method of measurement: these are *free ion conductivity* and *polarization conductivity*. The first is the ordinary conductivity due to the drift of free electrons or ions to the electrodes; the second is a conductivity determined by the energy dissipated as heat by the polarization currents in the dielectric. The latter bears the same relation to the neutral polarizable aggregates in the material, which carry the polarization currents, as does the free ion conductivity to the free ions in the dielectric. The terms free ion conductivity and polarization conductivity, or some other terms having approximately the same meaning, are essential to the discussion as they refer unambiguously to two distinct properties of the material, while the terms initial, final, true, infinite-frequency, a-c and d-c conductivity all refer to different methods of measuring these two properties of the material.

The current flowing in a dielectric to which a constant potential is applied often decreases with time for periods of the order of a few minutes or longer measured from the time of applying the potential. This decreasing current is variously referred to as a *residual charging current*, an *absorption current*, an *anomalous conduction current* or an *irreversible absorption current*, depending upon the interpretation given to the phenomenon. We have already indicated that these residual currents complicate the measuring technique in the determination of the d-c conductivity of insulating materials. The most definite kinds of residual currents are those which are simply a manifestation of the structural characteristics which give rise to anomalous dispersion of the dielectric constant. These residual currents and the residual charges associated with them will be referred to here as the *direct-current counterparts of anomalous dispersion* to indicate that they are not independent properties of the material, but necessary requirements of the existence of anomalous dispersion occurring at sufficiently low frequencies. The information obtainable from the study of such residual currents is the same in kind as that obtainable from the study of dielectric constant and conductivity by means of alternating currents; the residual phenomena, however, provide data regarding polarizations having relaxation-times which are too long for convenient investigation by alternating current methods. Residual currents of this kind have no significance in principle which is different from that of low-frequency a-c measurements.

Conductivity and Dielectric Loss

The conductivity of a material is usually thought of as a property which depends upon the ease with which electric charge can be transferred through the material by the application of an electric field, though it is recognized that a dissipation of electrical energy as heat occurs in the material through which the current is passing. In these terms we think of the conductivity as a quantity proportional to the current per unit voltage gradient, which in turn is proportional to the number of charge carriers, their mobility, and the magnitude of the charge borne by each carrier. For conductors it does not matter whether we define the conductivity, γ, as the factor by which the voltage gradient, E, must be multiplied to give the current density, I,

$$I = \gamma E \tag{1}$$

or as the factor by which the square of the voltage gradient must be multiplied to give the heat, W, developed per second in a unit cube of the material,[2]

$$W = IE = \gamma E^2, \tag{2}$$

for the heat developed by a given voltage is proportional to the current, no matter of what material the conductor is composed. This is due to the fact that the energy obtained by the moving charges from the applied electric field is dissipated continuously to the surrounding molecules or lattice structure as heat, and the electrons or ions then drift with constant average velocity in the direction of the applied field, developing heat at a rate proportional to the current.

However, the proportionality between current and heat developed which is characteristic of conductors does not obtain in dielectrics. When an alternating current flows in a dielectric it dissipates some electrical energy as heat; however, the amount is generally much smaller than would be dissipated by an equal current flowing in a

[2] Cf. for example, Mason and Weaver, "The Electromagnetic Field," Chicago (1929), p. 233.

conductor and, unlike conductors, the ratio of heat developed to current flowing varies with the material. This is due to the fact that most of the current flowing in a dielectric under ordinary conditions is a polarization current, or rather a sum of several polarization currents of different types, and in general a polarization current dissipates less energy as heat than an equal current flowing in a conductor. In fact a part of the current flowing in a dielectric—the optical polarization current—passes through the dielectric material without developing any heat in it at the ordinary frequencies of electrical transmission. Electrical energy can be transmitted through a good dielectric in a suitable range of frequencies with very little loss; in other words, the dielectric is transparent to currents which have a suitable frequency of alternation. In these circumstances the conductivity of the material as measured on a bridge would be very small though the current density per unit voltage gradient might be quite large. Evidently, then, the view of conductivity as simply a measure of the ease of transfer of electric charge through a material is not in general suitable for application to dielectrics.

The fact is that the *complex* conductivity represents the ease of displacement of electric charge in a dielectric while its real part (i.e., the a-c conductivity as measured on a bridge or equivalent measuring device) is the quantity to which the rate of heat development in the material is proportional. Therefore, in dealing with alternating currents flowing in dielectrics it is usually more convenient to regard the a-c conductivity as the factor which determines the rate of dissipation of electrical energy as heat in the material, rather than as a quantity which is proportional to the current density per unit voltage gradient or to the ease of displacement of electric charge in the material. In a later part of the discussion, however, it will be shown that the limiting high-frequency value of the a-c conductivity may be thought of as representing ease of displacement of electric charge, too, as in a conductor.

The heat developed in a dielectric by polarization currents is called *dielectric loss* and is analogous to the Joule heat developed by free electrons or ions in a conductor; however, it is a property of neutral aggregates of particles, such as polar molecules, rather than of free ions. In the case of a polarization due to polar molecules, for example, the equilibrium distribution of the orientations of the molecules is slightly changed by the application of an electric field. The dielectric constant depends upon the difference between the distribution of orientations with and without the applied field, while the dielectric loss represents the part of the energy of the applied field which is dissipated as heat because of the "friction" (i.e., the molecular equivalent of macroscopic friction) which the molecules experience as they change from the one equilibrium distribution of orientations to the other. Evidently, the dielectric loss may be quite as characteristic of the structure of the material as is the dielectric constant.

In an ideal insulating material there would be no free ion conduction, but in actual materials there are some free ions or electrons and these produce Joule heat as they drift towards the electrodes in the applied field. The total heat developed is the sum of the dielectric loss and the Joule heat; and, as the latter is proportional to the d-c or free ion conductivity, the dielectric loss is proportional to the total a-c conductivity (as measured on a bridge for example) less the d-c conductivity.

To give the discussion a more concrete basis, let us consider a dielectric which has a dielectric constant ϵ' and a loss-factor ϵ'' (or in other words which has a complex dielectric constant $\epsilon' - i\epsilon''$). Let it be contained in a parallel-plate condenser having a plate separation of d centimeters, and area A cm^2 for one surface of one of the plates. If a potential difference V is maintained between the plates of this condenser, a charge q per unit area will appear on either plate and a polarization P will be created in the dielectric. The current flowing in the leads to this condenser is Adq/dt, if we assume for the present that the conductivity due to free ions may be neglected. The conductivity is then given by

$$\gamma = \frac{1}{E}\frac{dq}{dt}, \tag{3}$$

where $E = V/d$. The charge q can be calculated from the dielectric constant of the material by means of relations which are provided by the general theory of electricity, namely

$$\epsilon E = D, \tag{4}$$

$$D = E + 4\pi P, \tag{5}$$

$$D = 4\pi q \text{ (for a parallel-plate condenser).} \tag{6}$$

So (3) becomes

$$\gamma E = \frac{dq}{dt} = \frac{1}{4\pi}\frac{dD}{dt} = \frac{\epsilon}{4\pi d}\frac{dV}{dt}, \tag{7}$$

where all of the electrical quantities are expressed in electrostatic units. When the applied potential is alternating, V may be expressed as the real part of $V = V_0 e^{i\omega t}$, where V_0 is the amplitude. The dielectric constant may then be written as the complex quantity $\epsilon' - i\epsilon''$, as shown in the preceding paper. The current density in the dielectric is then

$$\frac{dq}{dt} = i\omega(\epsilon' - i\epsilon'')\frac{V_0}{4\pi d}e^{i\omega t} \tag{8}$$

$$= \left(\frac{\epsilon''\omega}{4\pi} + i\frac{\epsilon'\omega}{4\pi}\right)E_0 e^{i\omega t} \tag{8a}$$

$$= (\gamma' + i\gamma'')E_0 e^{i\omega t}, \tag{8b}$$

where $\gamma' \equiv \epsilon''\omega/4\pi$ and $\gamma'' \equiv \epsilon'\omega/4\pi$. It is evident that $\gamma' + i\gamma''$ $(= \gamma)$ is the complex conductivity.

The dielectric constant and conductivity for alternating currents are determined by measurements, made with bridges or by other means, which give the admittance or impedance of the condenser containing the dielectric at the particular frequency at which the measurement is made. This admittance [3] may be expressed in terms of the equivalent parallel capacitance (C_p) and conductance (G_p) and an alternative expression for (8) is then

$$\frac{dq}{dt} = \frac{0.9 \times 10^{12}}{A}(G_p + iC_p\omega)V_0 e^{i\omega t}, \tag{9}$$

where G_p is expressed in mhos (or reciprocal ohms) and C_p in farads and 0.9×10^{12} is the ratio of the farad to the electrostatic unit of capacitance and also of the mho to the e.s.u. of conductance.

By comparing (9) with (8), (8*a*) and (8*b*) we obtain expressions for γ', ϵ'' and ϵ' in terms of the quantities C_p and G_p as directly measured on a bridge or similar arrangement. However, the expressions obtained are briefer if we make use of the fact that, when expressed in farads, the capacity C_0 of the empty condenser is

$$C_0 = \frac{A}{4\pi d \times 0.9 \times 10^{12}}. \tag{10}$$

Then it is evident that

$$\epsilon' = C_p/C_0, \tag{11}$$

$$\epsilon'' = G_p/C_0\omega, \tag{12}$$

$$\gamma' = G_p/4\pi C_0 \tag{13}$$

$$= \epsilon''\omega/4\pi = \epsilon'' f/2. \tag{13a}$$

[3] Measurements on a series bridge give directly the equivalent series resistance R_s and capacitance C_s. These data can be converted into equivalent parallel conductance and capacitance by the general relationships

$$C_p = \frac{C_s}{1 + (\omega R_s C_s)^2}; \qquad G_p = \frac{\omega^2 R_s C_s^2}{1 + (\omega R_s C_s)^2}.$$

In equations (11) to (13*a*) ϵ', ϵ'' and γ' are expressed in e.s.u., while C_p and C_0 are expressed in farads and G_p in mhos. The substitution of the frequency, f, for ω in (13*a*) depends upon the fact that $f = 2\pi\omega$.

While it is usual to express ϵ' and ϵ'' in e.s.u., it is more convenient for most purposes to have γ' in the units ordinarily used for specific conductance: thus when expressed in $\text{ohm}^{-1}\cdot\text{cm}^{-1}$

$$\gamma' = \frac{\epsilon''\omega}{4\pi \times 0.9 \times 10^{12}} = \frac{\epsilon'' f}{1.8 \times 10^{12}}, \tag{14}$$

$$= \frac{8.85 \times 10^{-2}}{C_{0\ \text{mmf}}} G_p = \frac{d}{A} G_p, \tag{14a}$$

where $C_{0\ \text{mmf}}$ is the capacitance in micromicrofarads.

By expressing equation (8) in the equivalent polar form certain quantities appear which are closely related to γ', ϵ' and ϵ'' and which are commonly used in describing the characteristics of dielectrics. The polar form is

$$\gamma = \gamma_0 e^{i\theta},$$

where $\gamma_0 = (\gamma'^2 + \gamma''^2)^{1/2}$, a quantity which is a measure of the amplitude of the complex current in the dielectric for unit voltage gradient, while $\theta = \tan^{-1} \gamma''/\gamma'$ is its phase angle. It is customary to use the *loss angle* which is defined as $\left(\frac{\pi}{2} - \theta\right) \equiv \delta$, rather than the phase angle in the description of dielectric properties. It is evident that $\delta = \tan^{-1} \gamma'/\gamma'' = \tan^{-1} \epsilon''/\epsilon'$ and that

$$\tan \delta = G_p/C_p\omega. \tag{15}$$

Similarly, the power factor is given by

$$\cos \theta = \gamma'/(\gamma'^2 + \gamma''^2)^{1/2}$$
$$= \epsilon''/(\epsilon'^2 + \epsilon''^2)^{1/2} = G_p/(G_p^2 + C_p^2\omega^2)^{1/2}. \tag{16}$$

When the current given by (8) is multiplied by the voltage, $E_0 \cos \omega t$, we obtain the instantaneous power, and from this the mean power $\overline{W}$ can be obtained by integration over a whole number of half periods. We then obtain

$$\overline{W} \text{ per second} = \gamma' \left(\frac{E_0}{\sqrt{2}}\right)^2 = \frac{\epsilon''\omega}{4\pi}\left(\frac{E_0}{\sqrt{2}}\right)^2 \tag{16a}$$

and

$$\overline{W} \text{ per cycle} = \frac{\epsilon''}{2}\left(\frac{E_0}{\sqrt{2}}\right)^2. \tag{16b}$$

This demonstrates the statements made earlier that γ' is proportional to the heat developed per second and ϵ'' to that developed per cycle in the dielectric. In the above equations $\bar{W}$ is in ergs per second or per cycle when E_0, γ', and ϵ'' are in e.s.u.

It can be seen from equation (8) that the total current flowing in the dielectric has a dissipative and a non-dissipative part: ϵ' is proportional to the non-dissipative part, and ϵ'' to the dissipative part. The loss-angle, ϵ''/ϵ', may be interpreted as the ratio of the dissipative to the non-dissipative current and the power factor as the ratio of the dissipative current to the total current.

The Frequency-Dependence of Conductivity

When the dielectric with which we are dealing possesses the property of anomalous dispersion, the expression for the loss factor ϵ'' as a function of frequency is

$$\epsilon'' = \frac{(\epsilon_0 - \epsilon_\infty)\omega\tau}{1 + \omega^2\tau^2}, \tag{17}$$

as was shown in the preceding paper. Substituting this expression for ϵ'' in (14) we obtain:

$$\gamma' = \frac{\epsilon''\omega}{4\pi} = \frac{1}{4\pi} \cdot \frac{(\epsilon_0 - \epsilon_\infty)\omega^2\tau}{1 + \omega^2\tau^2} \tag{18}$$

$$= \frac{1}{4\pi \times 0.9 \times 10^{12}} \cdot \frac{(\epsilon_0 - \epsilon_\infty)\omega^2\tau}{1 + \omega^2\tau^2}, \tag{18a}$$

where γ' is expressed in e.s.u. in (18) and in $\text{ohm}^{-1}\cdot\text{cm}^{-1}$ in (18*a*), and ϵ_0 is the static dielectric constant, ϵ_∞ the infinite-frequency dielectric constant and τ is the relaxation-time.

Differentiation of (18) with respect to frequency shows that γ' has no maximum when plotted against frequency; *the conductivity of any dielectric to which (18) applies should always increase with frequency, where it changes at all.* On the other hand, differentiation of (17) with respect to ω shows that the dielectric loss-factor has a maximum which occurs when $\omega\tau = 1$. The dielectric constant ϵ' is given by

$$\epsilon' = \epsilon_\infty + \frac{\epsilon_0 - \epsilon_\infty}{1 + \omega^2\tau^2} \tag{19}$$

and it will be seen that it shares with the conductivity the property of having no maximum when plotted against frequency. In Fig. 1 schematic curves are drawn which show the differences in the frequency dependence of γ', ϵ'' and ϵ' for a material having an absorptive polarization of relaxation-time τ. The conductivity goes up as the dielectric constant goes down, as if the one were being transformed into the other.

The most interesting feature of (18) is that as the frequency increases γ' approaches a limiting value, and that this limiting value,

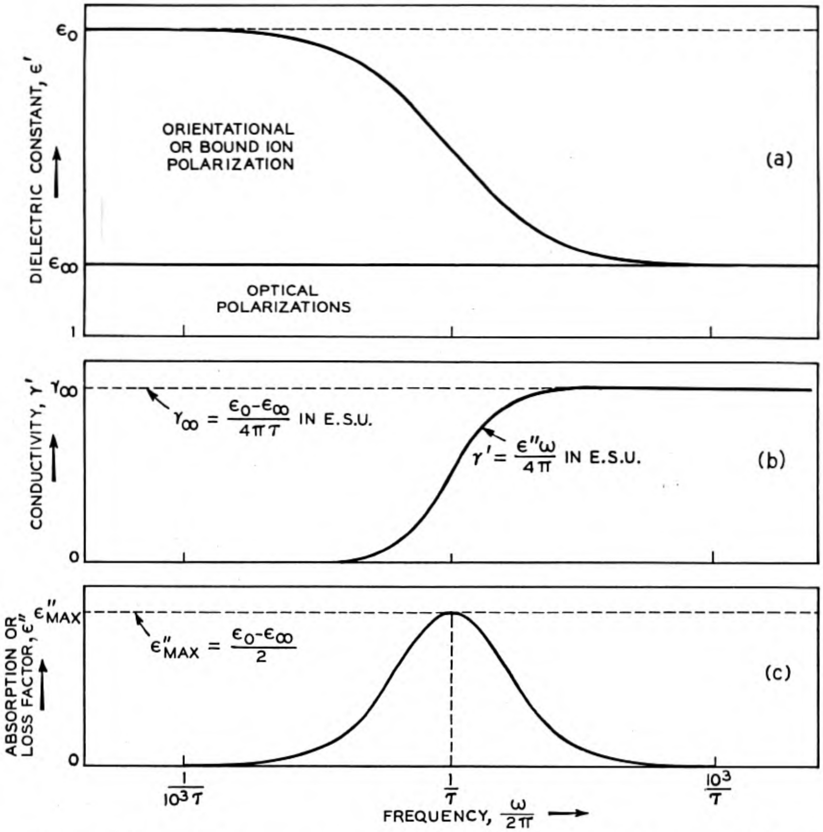


Fig. 1—Schematic diagram comparing the frequency dependence of dielectric constant (ϵ'), loss factor (ϵ'') and conductivity (γ'). This applies to a polarization having a single relaxation-time (τ). The frequency in cycles per second is $\omega/2\pi$.

which will be designated as γ_∞, has the value

$$\gamma_\infty = \frac{\epsilon_0 - \epsilon_\infty}{4\pi\tau} \tag{20}$$

$$= \frac{\epsilon_0 - \epsilon_\infty}{4\pi \times 0.9 \times 10^{12}\tau}, \tag{20a}$$

where (20) gives γ_∞ in e.s.u. and (20a) in $\text{ohm}^{-1}\cdot\text{cm}^{-1}$. The conthe dependence of conductivity on frequency in terms of the displacement and velocity of a single bound ion. As we are not concerned here with very high frequencies, we may employ the abbreviated equation of motion given in equation (17) of the preceding paper to

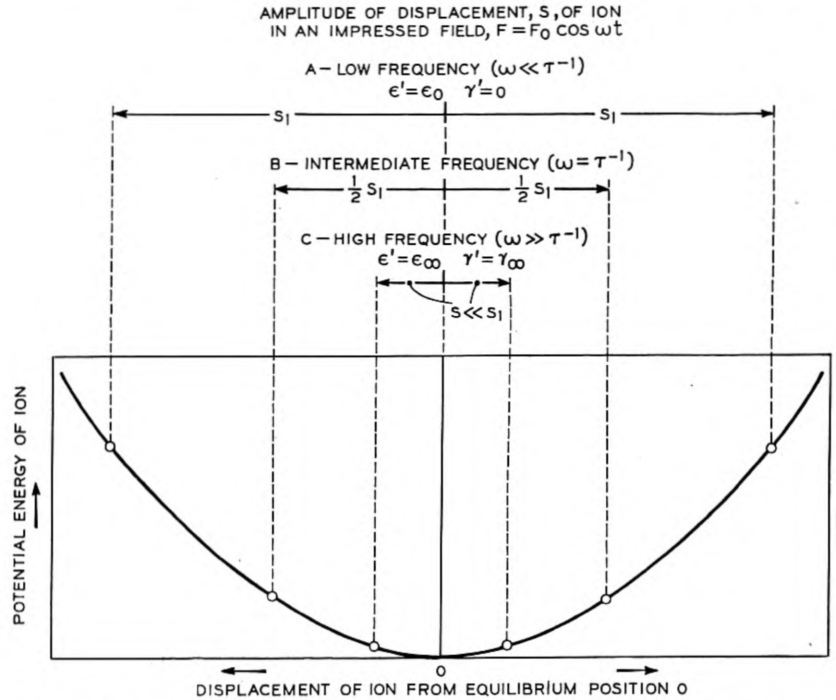


Fig. 3—The mechanism of anomalous dispersion illustrated by a simple model.

The model consists of a single bound ion. The potential energy of this ion increases when it is displaced from its equilibrium position 0. The ion also experiences a frictional force proportional to its velocity, as if it were an ion in solution. The upper part of the diagram shows the way in which the amplitude (s) of displacement of the ion by a given applied field varies with the frequency. It has its maximum amplitude at low frequencies (A in the diagram) and a comparatively negligible amplitude at high frequencies (C in the diagram). In this model the amplitude is a measure of the dielectric constant. The limiting value γ_∞ of the conductivity prevails under the conditions C where the amplitude is comparatively negligible.

discuss the motion of the bound ion of this model; this equation is

$$rv + fs = eF,$$

where $v = ds/dt$.

When a d-c voltage V_1 is applied to this model, it establishes a field F_1 which displaces the bound ion to a new equilibrium position s_1 if the field is allowed to act upon the ion for a sufficient time. The new equilibrium position s_1 of the ion corresponds to the static value P_1 of the polarization of the model. When the voltage is varying with the time according to $V = V_0 \cos \omega t$, the greatest amplitude which the displacement can have is s_1, and in general the amplitude will fall short of this value by an amount which increases with increasing frequency. The value s_1 is then closely approached only when the frequency is low as compared with the reciprocal of the relaxation-time, because at high frequencies the applied field reverses its direction before the ion has had time to reach s_1. At sufficiently low frequencies, namely where ω is negligible by comparison with $1/\tau$, the frictional dissipation of energy by the moving ion is so small that there is practically no difference between the instantaneous position of the ion when the voltage has any given value and the position it would finally attain upon reaching equilibrium for that voltage. The ion then moves through a succession of near-equilibrium positions, as in a reversible process in thermodynamics. The dielectric constant has its static value and the conductivity is zero unless there is a d-c conduction component in the total conductivity.

At the high-frequency extremity of a dispersion region we see that the situation is simply reversed: the alternations in the direction of the applied field are so rapid that the bound ion does not have time to move an appreciable distance from its equilibrium position before the direction of the applied field is reversed (C, Fig. 3). The amplitude of the displacement of the ion by the applied field is then small as compared with s_1 and the dielectric constant of the material receives practically no contribution from the bound ion of this model in these circumstances. However, though the amplitude of motion of the ion shrinks to a small fraction of s_1, its velocity is comparatively high and independent of frequency. The conductivity γ_∞ is proportional to the average velocity of the bound ion of the model under these conditions.

As the restoring force is proportional to the displacement, its effect upon the motion of the ion is negligible by comparison with that of the applied force when the displacement s is small as compared with s_1. On this basis, the fact that the conductivity is an increasing function of frequency may be attributed to the decrease in the influence of the restoring forces as the frequency increases. In fact, when the amplitude of displacement is very small as compared with s_1, the ion moves as if the only force opposing the applied force were the frictional force; that is, its average velocity is the same as that of a free ion subjected to the same applied field and the same friction.

For many of the purposes of this discussion we could use a model of the dielectric consisting of an air capacity C_s in series with a resistance

R_s both being shunted by a second air condenser C_∞. In this equivalent circuit, C_s and R_s refer to the polarizations responsible for anomalous dispersion, and C_∞ to the optical polarizations. The frequency-dependence of the equivalent parallel capacitance and conductance of this network is

$$C_p = C_\infty + \frac{C_0 - C_\infty}{1 + \omega^2 T^2} \tag{21}$$

and

$$G_p = \frac{(C_0 - C_\infty)\omega^2 T}{1 + \omega^2 T^2}, \tag{22}$$

where $C_0 \equiv C_s + C_\infty$ and $T \equiv C_s R_s$. In the above expressions $(C_0 - C_\infty)$ and T are analogous respectively to $\epsilon_0 - \epsilon_\infty$ and τ in equations (17) and (19).

The physical basis for the infinite frequency conductivity in this model depends upon the fact that at high frequencies the impedance of C_s is so low that nearly the whole drop in voltage is over the resistance R_s. This simple network is capable of representing the frequency-dependence of materials exhibiting anomalous dispersion due to a polarization having a single relaxation-time. In fact, when the frequency is sufficiently high that it is in the range where the conductivity is independent of frequency, the required network becomes even more simple, for it then reduces to C_∞ shunted by R_s, where the magnitude of C_∞ corresponds to ϵ_∞ and that of R_s to $1/\gamma_\infty$.

Polarization Conductivity

The operation of the models which have been discussed above provides a basis for interpreting the physical nature of γ_∞. The essential characteristics brought out by these models are listed below. They show the justification for considering γ_∞ to be a conductivity in the same sense as the ordinary d-c conductivity.

(1) To obtain γ' in an actual measurement on a dielectric, we subtract the d-c conductivity γ_f from the total a-c conductivity. There is then no contribution from free ion conduction in γ' and consequently none in γ_∞, its limiting value at high frequencies. Polar molecules or other polarizable aggregates in the dielectric must then be the origin of γ_∞.

(2) In the second place γ_∞ is independent of frequency, a property which puts it on the same footing as the d-c or free-ion conductivity in at least one respect.

(3) Earlier in this paper it was mentioned that the heat developed in a conductor for a given voltage is proportional to the *total* current,

but that in dielectrics this proportionality does not in general prevail. The current in a dielectric is complex and heat is developed only by its dissipative component. If the expressions for ϵ' and ϵ'' given in (17) and (19) are substituted in (8) we see that when ω becomes large by comparison with $1/\tau$, i.e., when γ' becomes γ_∞, the imaginary component of the current reduces to $\epsilon_\infty \omega/4\pi$; this is the optical polarization current. If it is subtracted from the total current given by (8), the remaining current contains no imaginary component. This current then develops as much heat in the dielectric as would a current of the same magnitude flowing in a conductor.

(4) In connection with the foregoing we see that unlike lower values of γ' the infinite-frequency conductivity γ_∞ is a measure of the ease with which electrical charge can be displaced in the material by a unit applied field. This characteristic of γ_∞ agrees with our usual conception of the physical basis of the conductivity of an electrolyte or a metal. (We assume in this connection that the optical polarization current $\epsilon_\infty \omega/4\pi$ may be neglected in comparison with the current responsible for γ_∞. Where this is not the case, appropriate modifications in the above statements are required.)

(5) It is characteristic of a dielectric that when the charged particles which form part of its structure are displaced by a force of external origin, there is a restoring force tending to return them to their initial positions. On the other hand, in an ideal conductor there are, by definition, no restoring forces of this kind. The above discussion of the model shows that in a dielectric possessing the property of anomalous dispersion it is possible to make the influence of the restoring forces on the motion of a bound ion negligible in comparison with that of the applied force by sufficiently increasing the frequency above the value corresponding to the reciprocal of the relaxation-time. This is the condition which prevails when γ' equals γ_∞. Thus at low frequencies ($\omega \ll \tau^{-1}$) the part of the dielectric structure which is responsible for anomalous dispersion behaves as a dielectric; whereas at high frequencies ($\omega \gg \tau^{-1}$) it behaves as a conductor. A result of this is that a dielectric exhibiting anomalous dispersion of the simple kind conforming to equation (39) of the preceding paper will behave in an electric circuit like a pure capacity shunted by a pure resistance over the whole of that range of frequencies where ϵ' and γ' are both practically independent of frequency and equal respectively to ϵ_∞ and γ_∞. Pure ice, for example, behaves in this manner over a considerable range of frequencies. (See Figs. 2 and 4.)

(6) The average velocity of the bound ion of our model becomes independent of frequency when ω is large as compared with $1/\tau$. This

constant velocity is equal to that which a free ion would have under the same voltage gradient if it were moving in a medium subjecting it to the same frictional resistance as is experienced by the bound ion of our model.

(7) In ordinary electrolytic conduction the conductivity is usually represented as the product of three factors: the number of ions per unit volume; their valence or charge per ion; and the average mobility of each ion, i.e., the average distance which an ion drifts per second in

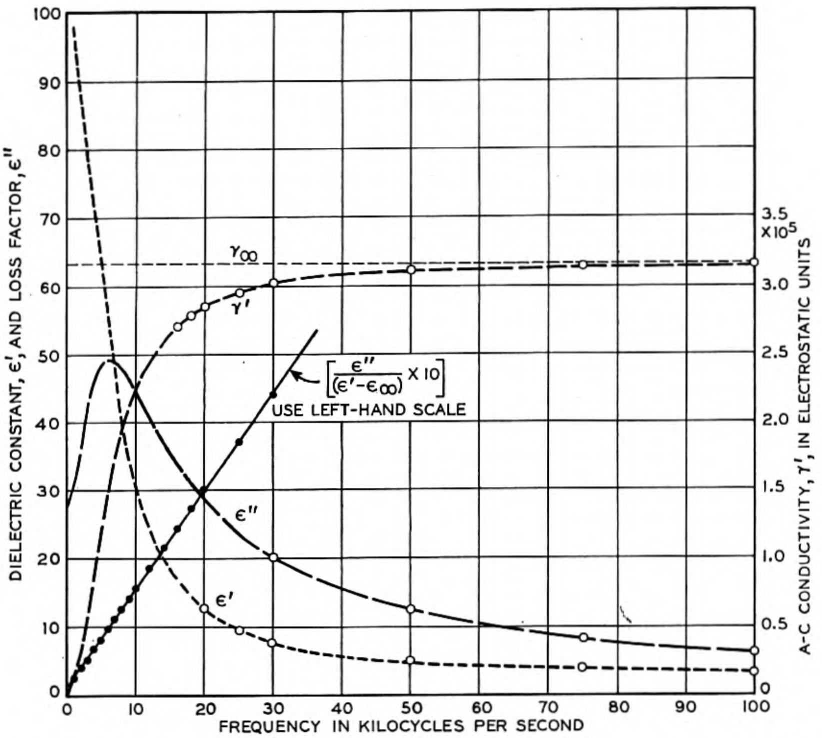


Fig. 4—Dependence of ϵ', ϵ'', γ' and $\epsilon''/(\epsilon' - \epsilon_\infty)$ upon frequency for ice at $-2.6°$ C.

the direction of the applied field. The above statement refers to one species of ion; when two or more species are present in the solution, the total current or conductivity is the sum of the currents or conductivities contributed by each type. The infinite-frequency conductivity γ_∞ may also be represented as the product of three factors similar in physical meaning to those just mentioned as applying to conduction by ions in solution. Thus reference to Table I, Item 2, will show that for the model we have employed to illustrate this discussion,

$$\gamma_\infty = \frac{1}{4\pi} \cdot \left(\frac{\epsilon_\infty + 2}{3} \right)^2 \cdot \frac{ne^2}{r} \cdot \tag{23}$$

The factors in this expression may be seen to have the following general significance: the quantity e is the charge on each bound ion; n is the number of bound ions per unit volume; and $e/4\pi r$ is a measure of the mobility per ion. This mobility does not refer to the motion of an ion which is free to move through the dielectric from one electrode to the other but to the mobility of a bound ion in small, local translational motions or the rotational mobility of a polar molecule. The remaining factor $\left(\frac{\epsilon_\infty + 2}{3} \right)^2$ is not of direct significance in the present connection.[5]

We see then that γ_∞ is also analogous to ordinary electrolytic conduction in that its physical mechanism may be represented as depending upon a mobility, a concentration and a factor such as dipole moment or charge per bound ion. The latter factor has a function in this mechanism which is similar to that of the valence or charge per ion in electrolytic conduction.

These considerations indicate that although γ_∞ is a property of polarizable units such as polar molecules it has the usual attributes of a conductivity due to free ions or free electrons. A dielectric which exhibits simple anomalous dispersion conforming to equations (17) and (19) then has *two* conductivities. One of these is the conductivity due to free ions; this will be called the *free ion conductivity* and designated throughout this paper by γ_f. The other is a conductivity which is a characteristic of the polarizable complexes responsible for anoma-

[5] When $\epsilon_\infty = 1$, equation (23) reduces to

$$\gamma_\infty = \frac{ne^2}{4\pi r},$$

showing thereby that the factor $\left(\frac{\epsilon_\infty + 2}{3} \right)^2$ would be absent if the material possessed no optical polarizations. Evidently γ_∞ depends upon the optical refractive index $\sqrt{\epsilon_\infty}$, as well as upon the characteristics of the absorptive polarization. In mixtures it may be possible to vary these two factors independently. Since optical polarization currents make no *direct* contribution to the energy dissipation in the dielectric, even up to the highest radio frequencies, it is interesting to observe that they make an *indirect* contribution according to equation (23). Their indirect action takes place by virtue of their effect on the actual internal field which acts upon each polarizable aggregate in the dielectric. The effect of the interaction of the optical polarization with the absorptive polarizations is to increase the apparent mobility of the polarizable complexes responsible for anomalous dispersion.

lous dispersion; this will be called the *polarization conductivity* [6] and designated by γ_{pol} in this paper.

The magnitude of the polarization conductivity of a material is proportional to the number of polarizable units of structure such as polar molecules or bound ions per unit volume which contribute to anomalous dispersion. It also depends upon the mobility which these polarizable units have in the local translational or rotational motions in which they engage in consequence of thermal agitation. It finally depends upon the permanent dipole moment of the polar molecules or upon the charges upon the bound ions.

The concentration of ions able to contribute to conduction in dielectrics is generally low because in many cases the free ion conductivity depends mainly upon a small percentage of impurity in the material. On the other hand, the concentration of polarizable units which are able to contribute to the polarization conductivity may be much larger and in fact even equal to the total number of molecules per unit volume. Consequently, the polarization conductivity γ_{pol} may often be more reproducible in measurements upon different specimens of the same dielectric than is the free ion conductivity.[7] It may well be that in many materials diffusion coefficients, thermal conductivity, mechanical dissipation and other similar properties which might be expected on theoretical grounds to be related to electrical conductivity will bear a simpler or more easily demonstrated relationship to polarization conductivity than to free ion conductivity.

An example of the advantage of using the infinite-frequency conductivity instead of the d-c conductivity appears in measurements of the conductivity of ice. In Fig. 5 the infinite-frequency conductivity (or polarization conductivity) of ice is plotted against the reciprocal of the absolute temperature, using unpublished data of the writers. The data for γ_∞ are reproducible and the curve shows a relation similar to that usually observed for the d-c conductivity of solids. Direct-current measurements on the same specimens on the other hand yielded very erratic results. It may be seen from Fig. 5 that the polarization conductivity is much higher than the d-c conductivity.

[6] We suggest for this conductivity the name polarization conductivity because it is a property of polarizable units of structure. In cases where the polarization is due to the change of orientation of polar molecules, we might instead refer to it as an *orientational conductivity* or a *polar molecule conductivity*, contrasting it thereby with the translational aspect of ordinary conduction by free ions. As ions which are loosely bound to some stationary or moving unit of the dielectric structure are often capable of producing anomalous dispersion, at least two types of polarization conductivity are possible; these may be described as the orientational conductivity and the bound ion conductivity.

[7] Joffé has obtained evidence that the initial conductivity, which we show here to be in some cases a polarization conductivity, is often superior in reproducibility to the final conductivity. (Cf. Reference 15.)

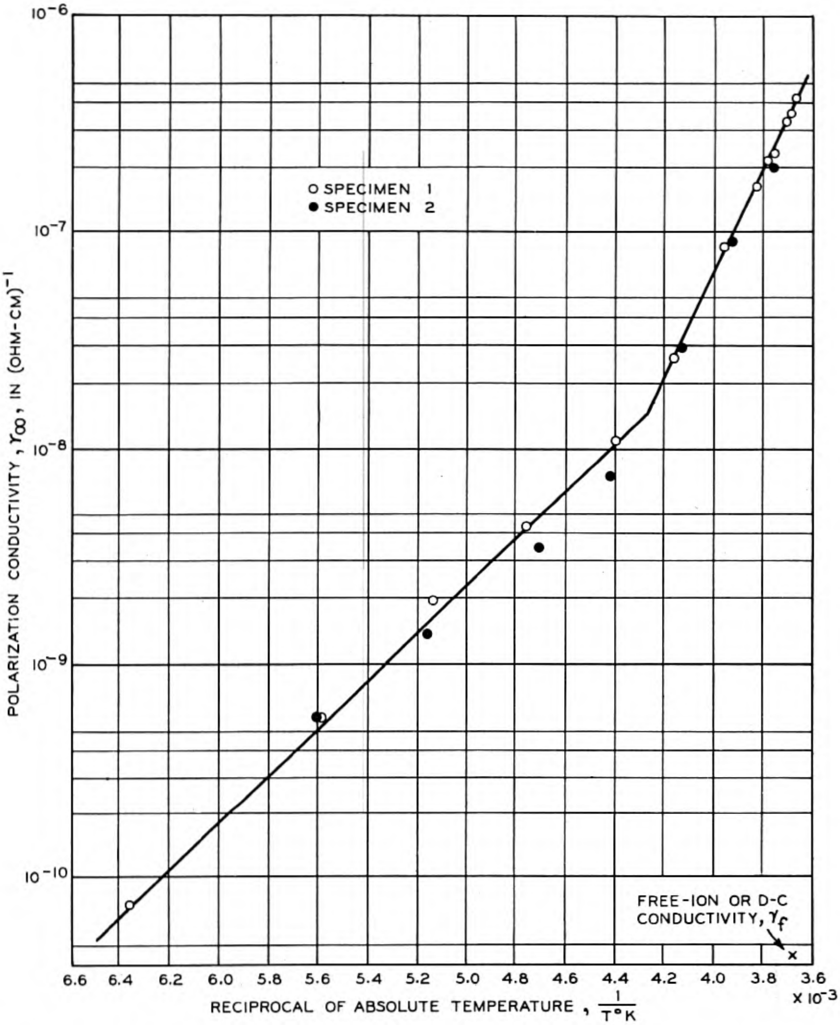


Fig. 5—An illustration of polarization conductivity. The temperature-dependence of the polarization conductivity of pure ice in the range − 0.8° C to − 190° C. The free-ion (or d-c) conductivity is also shown for a single temperature.

Types of Infinite-Frequency Conductivity

In Table I of the preceding paper there were listed several types of polarization capable of yielding anomalous dispersion curves distinguishable from each other only by the values of the constants. Corresponding to each is a different expression for the polarization conductivity and these expressions are listed in Table I of this paper.

This list shows that there are at least three main types of infinite-frequency conductivity. The first of these is the type which depends upon the change of the orientation of polar molecules according to the Debye theory. This type of polarization conductivity is of more theoretical interest than any of the others and perhaps also of more practical importance; as already mentioned, it may be described as an orientational conductivity to emphasize that no translational mobility is necessary for it to occur.

TABLE I

EXPRESSIONS FOR THE CONSTANT VALUE γ_∞ APPROACHED BY THE A–C CONDUCTIVITY AS THE FREQUENCY INCREASES

Type of Polarization	
1. The Orientational Polarization due to Polar Molecules	$\gamma_\infty = \frac{L\mu^2}{12\pi\eta a^3}\cdot\left(\frac{\epsilon_\infty + 2}{3}\right)^2$
2. A Distortional Polarization having a Relaxation-Time given by $\tau = r/f$	$\gamma_\infty = \frac{ne^2}{4\pi r}\cdot\left(\frac{\epsilon_\infty + 2}{3}\right)^2$
3. Polarizations due to Spatial Variations of Conductivity and Dielectric Constant	
(*a*) A two-layer dielectric, layers 1 and 2 having respectively static dielectric constants ϵ_1 and ϵ_2 and free-ion conductivities γ_1 and γ_2	$\gamma_\infty = \frac{(\epsilon_1\gamma_2 - \epsilon_2\gamma_1)^2}{(\epsilon_1 + \epsilon_2)^2(\gamma_1 + \gamma_2)}$
(*b*) Special case of (3*a*) where γ_1 is much larger than γ_2	$\gamma_\infty = \left(\frac{\epsilon_2}{\epsilon_1 + \epsilon_2}\right)^2\gamma_1$
(*c*) Special case of (3*b*) where $\epsilon_1 \cong \epsilon_2$	$\gamma_\infty = \gamma_1/4$
(*d*) Special case of (3*a*) consisting of a high resistance transition layer at the dielectric/electrode interface, where γ_1 is the conductivity of the dielectric	$\gamma_\infty = \gamma_1$
(*e*) Conducting spheres dispersed in an insulating medium of the same dielectric constant	$\gamma_\infty = p\gamma_1$

Note: The infinite-frequency conductivity γ_∞ is given here in e.s.u. (See equation 20.) In Table I of the preceding paper (*B. S. T. J.*, *17*, 640 (1938)) the values of $(\epsilon_0 - \epsilon_\infty)$ and τ are given for the polarizations listed above; the expressions given there for τ should be divided by 4π in the case of Items 3*a*, *c*, *d* and *e*, as should also the expressions for $\epsilon_0 - \epsilon_\infty$ in the case of Item 2. The quantities which appear in the above Table are defined in an appendix to the preceding paper. γ_1 and γ_2 are expressed in e.s.u. in this table.

The remaining members of the list of Table I originate in the properties of ions rather than those of molecules. These ions must be more or less bound in order to have an infinite-frequency conductivity differing from the zero-frequency conductivity.[8] The nature and

[8] It will be recalled that the terms infinite-frequency and zero-frequency are not used here in their general meaning but merely as a convenient way to indicate opposite directions of extrapolation of dispersion curves. They refer respectively to the high-frequency extremity of a dispersion curve where γ' becomes practically independent of frequency and to the low-frequency extremity where the dielectric constant becomes practically independent of frequency.

strength of the binding forces may vary widely amongst the types of polarization which have an infinite-frequency conductivity. An ion will be regarded as bound if its potential energy increases when it is displaced from an equilibrium position by an applied field.[9]

Included among the infinite-frequency conductivities which depend upon the presence of ions in the dielectric is a type for which macroscopic inhomogeneities in the dielectric are responsible; for example, a two-layer or multiple layer laminated dielectric, or a dielectric in which space-charges [10] form because of spatial variations in its resistivity or because of a transition layer of high resistance at the contact between dielectric and electrode. Examples of the infinite-frequency conductivity due to this type of mechanism are given in Items 3*a*, *b*, *c* and *d* of Table I.

This type of infinite-frequency conductivity is of little interest in principle, but in practice there may be many instances in which the measurement of the infinite-frequency conductivity provides a convenient means of determining the conductivities of the constituents of these non-homogeneous systems. For example, when one layer of a two-layer dielectric has a much higher conductivity than the other, γ_∞ assumes a value which is related simply to the free ion conductivity of the layer which has the higher conductivity (see Item 3*b*, Table I). If the dielectric constants of the two layers are equal, γ_∞ is equal to one-quarter of the conductivity of the high-conductivity layer. The conductivities γ_1 and γ_2 of the two layers are considered in the present connection to be free ion conductivities. A more complicated situation is possible where γ_1 and γ_2 are in part polarization conductivities due to polar molecules or bound ions.

The special case of a space-charge caused by a thin layer of high resistance at one or both of the electrodes is of interest in connection with the methods recommended by Joffé for the measurement of the true conductivity of crystals. This will be discussed in more detail later but for the present it may be noted that the infinite-frequency

[9] It is necessary to confine the application of the last statement to direct voltages or to frequencies lower than those for which γ' is equal to γ_∞. When the frequency is high enough for the latter condition to prevail the amplitude of displacement of the ion becomes so small that the applied field produces no appreciable increase in the average potential energy of the ion; this is illustrated in Fig. 3 at *C*.

[10] The external effects of a space-charge occurring in a dielectric because of spatial variations in its resistivity may be reproduced by a uniformly distributed polarization of suitably adjusted magnitude and relaxation-time. Several different polarizations of different magnitudes and relaxation-times would in some instances be required. In referring to such a space-charge as a polarization we may think of the term as applying to the uniform distribution of polarization which could replace the space-charge in its external effects.

conductivity for this system has the simple value

$$\gamma_\infty = 1/4\pi C_0 R = \gamma_1$$

as shown in Item 3*d*, Table I. Here γ_1 is the free ion conductivity of the main part of the dielectric, R is its resistance and $4\pi C_0(= A/d)$ is the ratio of thickness to length of specimen. (In Table I of the preceding paper the symbol C_∞ is used in place of C_0.) *The infinite-frequency conductivity in a non-homogeneous system of this type is a free-ion conductivity.*

Bound ions may also be distributed with macroscopic uniformity in a dielectric. An example of this type of bound ion conductivity [11] is one due to conducting particles dispersed uniformly in a relatively non-conducting medium. This is the case referred to in Item 3*e* of Table I. Macroscopic uniformity is obtained in this case by the random distribution of a large number of particles. However, there are some general experimental indications that the distribution of bound ions may in some materials depend upon the basic internal structure of the dielectric and involve some regular geometrical configuration repeated throughout the material.

In certain dielectrics which absorb an appreciable amount of water when in a humid atmosphere, conduction takes place in aqueous conduction paths permeating the solid. Examples of these materials are cotton, paper, silk and wool. This property is probably shared by many other polymeric substances. The water in these materials is distributed in minute capillaries, the dimensions and other characteristics of which probably determine the form and distribution of the conduction paths.[12] There exists in these materials a condition capable of producing a bound ion conductivity inasmuch as there are indications that the conducting paths are not of uniform cross-sectional area. Evidence for the existence of a bound ion conductivity in the kind of material to which we have just referred is provided, for example, by conductivity measurements on cotton.[13] Raw cotton contains salts which can be removed by extraction leaving the material otherwise practically unchanged. These salts are likely to be distributed in the material with macroscopic uniformity as they form part of its natural structure. The fact that the removal of these salts decreases the

[11] As already mentioned we shall call any infinite-frequency conductivity which is caused by a macroscopically uniform distribution of bound ions a *bound ion conductivity*. In some places this term will be applied to any conductivity due to bound ions irrespective of whether or not that conductivity is the limiting high-frequency value.

[12] One of the basic structural units of cellulose and other similar materials is the *micelle*. This usually contains a large number of molecules and the capillaries we refer to may correspond to the intermicellar spaces.

[13] Murphy, *Journal of Physical Chemistry*, *33*, 200 (1929).

dielectric loss indicates that the material possessed a bound ion conductivity before the salts were removed.

Some of the materials belonging to the class of dielectrics which we have just discussed are closely related in chemical and in physical structure to compounds which are important biologically and in the study of plant and seed structure. Many are also of commercial importance as insulating materials.

It is not necessary that the conduction paths be composed of aqueous solutions: in some materials plasticizers or products of pyrolysis are sufficiently conducting for this purpose. The dielectric behavior of certain plastics may be interpreted as evidence for the existence of such non-aqueous conduction paths in the material, producing a free ion conductivity, a bound ion conductivity and a contribution to the dielectric constant. Imperfections of structure occurring in crystals are able to produce a bound ion conductivity and there is experimental evidence that these imperfections do occur.[14] The regular lattice ions in an ionic crystal have too high a binding energy, and dissipate too little energy in their motions in a radio frequency electric field to produce a bound ion conductivity.

The polarization which is responsible for the bound ion conductivity is of the interfacial, or Maxwell-Wagner, type. This type of polarization may be of importance in materials with a cellular structure and in materials which may be described as interstitially conducting dielectrics.

In the above discussion we have outlined the character of three widely different types of infinite-frequency conductivity:

(*a*) An orientational conductivity depending upon the small changes which an applied field produces in the average orientation of polar molecules.

(*b*) A bound ion conductivity depending upon the displacement of uniformly distributed bound ions.

(*c*) An infinite-frequency conductivity which is proportional to the free ion conductivity of one of the constituents of a dielectric consisting of two or more layers of widely different conductivities.

THE RELAXATION-TIME

The relaxation-time is closely related to the infinite-frequency conductivity. This may be seen by reference to equation (20), which shows that the relaxation-time is given by

$$\tau = \frac{\epsilon_0 - \epsilon_\infty}{4\pi\gamma_\infty}. \tag{24}$$

[14] See, for example, A. Smekal, *Zeits. f. techn. Physik*, *8*, 561 (1927).

This equation shows that specifying the values of τ and $(\epsilon_0 - \epsilon_\infty)$ gives as much information as specifying γ_∞ and $(\epsilon_0 - \epsilon_\infty)$. In some applications there are advantages in using τ but in other applications greater simplicity of description is gained by using γ_∞.

There are several convenient ways of calculating the relaxation-time. The more familiar ones depend upon the position of maxima which occur in certain dielectric properties when they are plotted against the frequency: there are maxima in the loss factor vs. frequency curve, in the tangent of the loss angle vs. frequency curve and in the power factor vs. frequency curve. As these maxima occur at different frequencies, the corresponding expressions for the relaxation-time are also different. They are listed in Table II. It will be observed that the simplest of these formulae for the relaxation-time is the one involving the maximum in the loss factor.

TABLE II

List of Formulae for Calculating the Relaxation-Time (τ)

1. The frequency at which the maximum in loss factor (ϵ'' or $\epsilon' \tan\delta$) occurs is $\omega_{max(1)}$	$\tau = 1/\omega_{max(1)}$
2. The frequency at which the maximum in loss angle (ϵ''/ϵ' or $\tan\delta$) occurs is $\omega_{max(2)}$	$\tau = \sqrt{\frac{\epsilon_\infty}{\epsilon_0}}\frac{1}{\omega_{max(2)}}$
3. The frequency at which the maximum in power factor $\epsilon''/(\epsilon'^2 + \epsilon''^2)^{1/2}$ occurs is $\omega_{max(3)}$	$\tau = \sqrt{2}\sqrt{\frac{\epsilon_\infty}{\epsilon_0}}\frac{1}{\omega_{max(3)}}$
4. The quantity $\epsilon''/(\epsilon' - \epsilon_\infty)$ is a linear function of ω	$\tau = \frac{d}{d\omega}(\epsilon''/(\epsilon' - \epsilon_\infty))$
5. The relaxation-time is proportional to the ratio of the absorptive part $(\epsilon_0 - \epsilon_\infty)$ of the static dielectric constant to the infinite-frequency conductivity (γ_∞)	$\tau = (\epsilon_0 - \epsilon_\infty)/4\pi\gamma_\infty$

Note: An example of the application of these formulae is provided by the curves of Fig. 4. The value of τ for ice at $-2.6°C$ is 25.8 microseconds as calculated from the position of the maximum in ϵ'', 24.6 microseconds as calculated from $(\epsilon_0 - \epsilon_\infty)/4\pi\gamma_\infty$, where γ_∞ is in e.s.u., and 23.1 microseconds as calculated from the slope of $\epsilon''/(\epsilon' - \epsilon_\infty)$.

The function $\epsilon''/(\epsilon' - \epsilon_\infty)$ is a linear function of ω with slope equal to τ. This property provides an alternative method of calculating the relaxation-time. An example of its application to an actual material is provided by the data for ice plotted in Fig. 4.

It is interesting that $\epsilon''/(\epsilon' - \epsilon_\infty)$ has no maximum while ϵ''/ϵ' has a maximum. The physical basis of this is that in subtracting ϵ_∞ from ϵ' we remove the contribution to ϵ' made by optical polarizations. What is left represents only the dielectric constant due to the polarization responsible for anomalous dispersion. Consequently the tangent of the loss angle of the polarization current responsible for anomalous dispersion has no maximum when plotted against the frequency and its behavior is, therefore, in contrast with the tangent of the loss angle of the *total* polarization current, i.e., the sum of the optical polarization current and the absorptive polarization current.

From the above discussion it will also be evident that the physical basis for the maximum in tan δ is different from that of the maximum in ϵ''. As we have just shown, the maximum in ϵ''/ϵ' (= tan δ) depends upon the inclusion of optical polarizations in ϵ'. On the other hand, there would be a maximum in ϵ'' even if the part of the dielectric constant which is due to optical polarizations (ϵ_∞) were neglected or considered to be zero. This will be evident by differentiation of equation (17).

The maximum in ϵ'' is an intrinsic property of the absorptive polarization. The general nature of the mechanism by which it occurs is as follows: ϵ'' is proportional to γ'/ω; as the frequency increases γ'/ω at first increases, but when γ' reaches the constant value γ_∞, further increase in frequency causes γ'/ω to decrease.

The quantity τ which we have discussed here is a property of the dielectric as a whole as we indicated in the preceding paper. This quantity is connected with the relaxation-time τ' of the individual polarizable units by the relation

$$\tau = \frac{\epsilon_0 - 2}{\epsilon_\infty + 2}\tau', \tag{25}$$

when the material is of cubic or isotropic structure. This relationship is a consequence of the fact that the actual force acting upon a particle within a dielectric depends not only upon the applied field of external origin but also upon a force exerted by the polarization induced in the dielectric.

The Relationship between Dielectric Constant and Dielectric Loss

If $\epsilon'_{\max}$ is the value of the dielectric constant at the frequency where the loss factor ϵ'' is at a maximum when plotted against frequency, we have

$$\epsilon'_{\max} = \epsilon_\infty + \frac{\epsilon_0 - \epsilon_\infty}{1 + \omega^2_{\max}\tau^2} = \frac{\epsilon_0 + \epsilon_\infty}{2}, \tag{26}$$

$$\epsilon''_{\max} = \frac{(\epsilon_0 - \epsilon_\infty)\omega_{\max}\tau}{1 + \omega^2_{\max}\tau^2} = \frac{\epsilon_0 - \epsilon_\infty}{2}. \tag{27}$$

By addition and subtraction of (26) and (27) we obtain the following relationships:

$$\epsilon_0 = \epsilon'_{\max} + \epsilon''_{\max}, \tag{28}$$

$$\epsilon_\infty = \epsilon'_{\max} - \epsilon''_{\max}, \tag{29}$$

$$\epsilon_0 - \epsilon_\infty = 2\epsilon''_{\max}. \tag{30}$$

The comparison of the last equation with equation (20) brings out an interesting contrast between the maximum dielectric loss per *cycle* ($\epsilon''_{\max}$) and the maximum dielectric loss per *second* (γ_∞). *The maximum dielectric loss per cycle is completely determined by the difference between the static dielectric constant and the optical dielectric constant.* On the other hand, the dielectric loss per second depends as well upon the relaxation-time. The relaxation-time usually varies rapidly with temperature, whereas ($\epsilon_0 - \epsilon_\infty$) changes comparatively slowly with temperature. The temperature-dependence of the maximum dielectric loss per cycle is related to the polarizability of the material, whereas the temperature variation of the maximum dielectric loss per second is primarily a measure of the change of internal friction with temperature.

In the foregoing we have discussed the conductivity, dielectric loss and relaxation-time of dielectrics which have simple properties with respect to the frequency-dependence of these quantities. However, for many dielectrics, particularly solids, the experimental data are not in agreement with the dispersion formulae for a single relaxation-time which has been discussed here. The explanation usually adopted for this discrepancy is that the polarizations induced in the dielectric possess a distribution of relaxation-times.

The D-C Counterparts of Anomalous Dispersion

A dielectric so constructed that it exhibits anomalous dispersion under an alternating voltage should show some equally characteristic behavior when a direct voltage is substituted for the alternating one. These characteristics may be described as the d-c counterparts of anomalous dispersion. They include certain definite types of variation of current with time under constant applied potential.

In the appendix the d-c counterparts of anomalous dispersion are derived by employing the model used throughout this paper. This enables us to demonstrate an especially simple relationship between the a-c and d-c conductivity.

Equation (16) of the appendix gives the apparent conductivity as a function of charging time. As it has been assumed that $\epsilon_\infty C_0 R \ll \tau$, the first term of equation (16) will quickly become negligible. Then for charging times, measured from the instant of applying the voltage, such that $\tau \gg t_c \gg \epsilon_\infty C_0 R$, the apparent conductivity $\gamma_c(t_c)$ has the value

$$\gamma_0 = \frac{\epsilon_0 - \epsilon_\infty}{4\pi\tau} = \gamma_\infty. \tag{31}$$

The special value of $\gamma_c(t_c)$ which we have designated as γ_0 in (31) will be called the *initial conductivity.*

Equation (31) states that the infinite-frequency conductivity (γ_∞), *obtained by a-c measurements, is equal to the initial d-c conductivity* (γ_0) *which would be obtained by extrapolating current-time curves toward the instant of applying the voltage.* This relationship, which has not been demonstrated previously to our knowledge, is of interest in connection with the interpretation of conductivity measurements.

The model described in Appendix I and indicated schematically in Fig. 3 illustrates the physical nature of the initial conductivity in a simple manner. When a constant voltage V_1 is applied to a dielectric having the properties of this model an effective impressed field F_1 is established in the dielectric. This displaces the bound ion assumed to be responsible for the polarization in this model. The magnitude of the displacement s of this bound ion depends upon the length of time that F_1 is applied. If it is applied for a time which is much longer than the relaxation-time, s approaches a constant value s_1 corresponding to complete polarization of the dielectric. On the other hand, during that stage of the charging process when the charging time t_c is negligible in comparison with the relaxation time τ and at the same time large as compared with the time constant $\epsilon_\infty C_0 R$, the displacement s is negligible in comparison with s_1. The resultant force tending to displace the ion is then approximately equal to the impressed force; $eF_1 - fs_1 \cong eF_1$.

Near the beginning of the charging process there is a brief interval of time ($\epsilon_\infty C_0 R \ll t_c \ll \tau$), when the motion of the bound ion of our model in the applied field is essentially the same as that of a free ion. During this interval the prevailing conductivity is the initial conductivity defined by equation (31). Although the bound ion of the model is actually subjected to a force tending to restore it to its initial position, this restoring force has not had time during the initial stage of the charging process to build up to a magnitude appreciable in comparison with the applied force. The initial conductivity corresponds to a condition in the dielectric where bound ions act for a brief time as if they were free as far as conduction processes are concerned. The infinite-frequency conductivity corresponds also to this condition. The variation of apparent conductivity with the time of charging is indicated schematically in Fig. 6. As is there shown, there must be an initial stage, always too brief to be detected experimentally, during which the inductance of the circuit and the inertia of the charges cannot be neglected.

The Types of Initial Conductivity

The polarization conductivity γ_{pol} may then be measured in two ways: either as the infinite-frequency conductivity γ_∞ obtained by a-c measurements, or as the initial conductvity γ_0 obtained by d-c measurements:

$$\gamma_{\text{pol}} = \gamma_\infty = \gamma_0.$$

The quantity γ_{pol} refers to a property of the material, whereas γ_0 and γ_∞ refer to the methods of measuring this property.

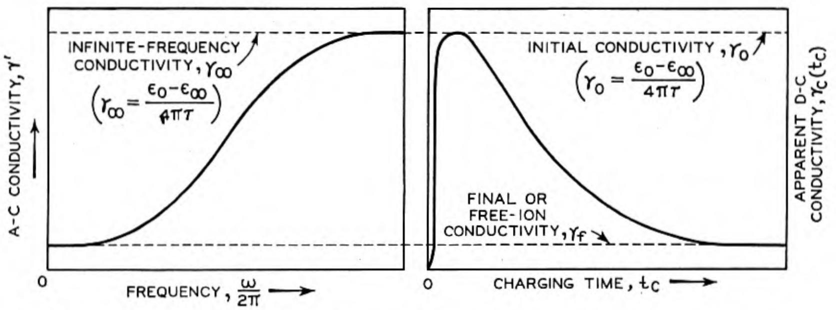


Fig. 6—A-c and d-c methods of measuring conductivity. A schematic diagram comparing the dependence of a-c conductivity (γ') on frequency with the dependence of apparent d-c conductivity $\gamma_c(t_c)$ on charging time (t_c). For homogeneous dielectrics $\gamma_0 = \gamma_\infty = \gamma_{\text{pol}}$, where γ_{pol} is a conductivity due to polar molecules or to bound ions. For some non-homogeneous dielectrics $\gamma_0 = \gamma_\infty = \gamma_f$, where γ_f is a conductivity due to free ions.

Reference to Table I will show that there are several types of polarization which may be responsible for γ_∞. Consequently this will be true also of the initial conductivity. The equality of γ_0 and γ_∞ applies to all of these types of polarization and, therefore, can not be used to distinguish between them. For this reason experimental agreement between a-c and d-c measurements of conductivity does not enable us to distinguish whether we are dealing with polar molecules, bound ions, or free ions in a macroscopically inhomogeneous dielectric. Thus it is clear that when we are dealing with a polarization due to polar molecules, the *initial conductivity is a true polarization conductivity or a specific dielectric loss.* If the polarization is of the bound ion type discussed earlier in this paper, the initial conductivity is also a polarization conductivity. However, if we are dealing with a macroscopically non-homogeneous dielectric such as a two-layer dielectric or a material *which has a high resistance blocking layer at one of the electrodes* γ_0 *is a free ion conductivity.* This is evident from the previous discussion of the significance of γ_∞.

These relationships are of interest in connection with the difficulties encountered in the interpretation of d-c conductivity data which were mentioned in the introduction. They apply especially to the methods recommended by Joffé [15] and by Richardson.[16]

When a constant potential difference is maintained between the plates of a condenser containing a solid dielectric the current observed does not in general remain constant but usually decreases with time. This decrease may continue for several minutes or hours. The study of these residual currents is of importance in connection with the interpretation of conductivity data on dielectrics. The question arises as to how much of the observed behavior of the residual currents which flow in dielectrics under constant potential may be explained as the d-c counterparts of anomalous dispersion. Many materials of practical importance as insulators exhibit a more complicated type of variation with frequency than is indicated by equations (17) and (19) which refer only to the simplest observed type of dispersion. The more complicated types of behavior observed are usually attributed to the presence of polarizations possessing a wide distribution of relaxation-times. However, other processes may also contribute to the deviation of the experimental curves from the theoretical. One of these is electrolysis which produces changes in the composition, and consequently in the conductivity of the material. Another effect which may contribute is a possible lack of constancy of the relaxation-time. It is evident, therefore, that the d-c counterparts of anomalous dispersion due to a polarization of a single relaxation-time should not be expected to explain all of the observed residual phenomena, particularly in solid dielectrics. However, the quantitative relations derived as the d-c counterparts of anomalous dispersion are applicable to some materials, and for those to which they are not quantitatively applicable, may serve as a useful guide in the interpretation of the residual currents. As another section of this paper is planned in which the influence of residual currents upon conductivity measurements will be discussed further, we have included in the appendix some relationships which are useful in the interpretation of the behavior of these currents.

[15] A. Joffé, *Ann. d. Physik*, *72*, 481 (1923); "Physics of Crystals," New York (1928); *Zeits. f. Physik*, *62*, 730 (1930). See also Sinjelnikoff and Walther, *Zeits. f. Physik*, *40*, 786 (1927).

[16] S. W. Richardson, *Proc. Roy. Soc.*, *107A*, 102 (1925).

APPENDIX

The D-C Counterparts of Anomalous Dispersion

Having developed the a-c characteristics of the model, the properties of which were specified in the preceding paper,[17] we now turn to the direct-current characteristics of this model. This involves investigating the characteristics of the currents produced when a constant or direct voltage V_1 is applied to a condenser containing a dielectric having properties which correspond in all respects essential to the discussion to those of the model just described. Let the resistance of the leads to the condenser be R and let the source of the electromotive force V_1 have a negligible internal resistance. These conditions require the following five equations to be satisfied simultaneously:[18]

$$r\frac{dP}{dt} + fP - ne^2F = 0, \tag{1}$$

$$F = E + AP_t, \tag{2}$$

$$P_t = k_iF + P, \tag{3}$$

$$D = E + P_t, \tag{4}$$

$$E = E_1 - CR\frac{dD}{dt}. \tag{5}$$

(Rational e.s.u. are used in these equations for convenience but the final relations are converted into electrostatic units.[19]) P_t denotes the total polarization; it is the sum of the optical polarizations, given by k_iF, and the polarization P which forms comparatively slowly and causes anomalous dispersion. The total displacement D, is given by (4). C_0 is the air capacitance of the condenser. V is the potential drop over the condenser. The separation of the plates of the condenser is d. When the whole drop in potential is concentrated over the condenser the applied field strength E_1 is given by $E_1 = V_1/d$. Equation (5) is obtained by equating the total drop in potential over the leads and the condenser to the applied potential V_1.

Equations (1) to (5) may be combined to give the following:

$$(1 + 2p_0)\tau' CR\frac{d^2D}{dt^2} + \{(1 - p_0)\tau' + CR(1 + 2p_0 + 2p_1)\}\frac{dD}{dt} + (1 - p_0 - p_1)D - (1 + 2p_0 + 2p_1)E_1 = 0. \tag{6}$$

[17] These properties are also outlined on page 513 of this paper.

[18] Cf. P. Debye, "Polar Molecules," pp. 86–88, where an analogous case is discussed.

[19] For conversion factors see, for example, Mason and Weaver, Reference 2, page 370. The rational electrostatic unit of conductivity is smaller than the e.s.u.; the ratio is 4π. The dielectric constant is unaffected by changing from rational e.s.u. to e.s.u.

In this equation the following abbreviations are used: $\tau' = r/f$, $Ak_i = p_0$, $Ak = p_1$. The last two abbreviations are introduced to facilitate comparison with an analogous derivation given in "Polar Molecules," p. 86. $D = D_0e^{\alpha t}$ is a solution of the homogeneous differential equation obtained by letting $E_1 = 0$, provided that the following equation is satisfied:

$$\alpha^2 + \left\{ \left(\frac{1 - p_0}{1 + 2p_0} \right) \frac{1}{CR} + \frac{(1 + 2p_0 + 2p_1)}{1 + 2p_0} \frac{1}{\tau'} \right\} \alpha + \left(\frac{1 - p_0 - p_1}{1 + 2p_1} \right) \frac{1}{\tau' CR} = 0. \quad (7)$$

As we are interested here only in the special case where $CR \ll \tau'$, the $1/\tau'$ term in the coefficient of α can be neglected in comparison with the $1/CR$ term, and the roots are

$$\alpha = -\left(\frac{1 - p_0 - p_1}{1 - p_0} \right) \frac{1}{\tau'} \text{ or } \alpha = -\left(\frac{1 - p_0}{1 + 2p_0} \right) \frac{1}{CR}$$

to a degree of approximation which improves the larger the ratio τ/CR. The general solution of the non-homogeneous equation (6) is

$$D = D_1 e^{-\left(\frac{1-p_0}{1+2p_0}\right)\frac{t}{CR}} + D_2 e^{-\left(\frac{1-p_0-p_1}{1-p_0}\right)\frac{t}{\tau'}} + \left(\frac{1 + 2p_0 + 2p_1}{1 - p_0 - p_1} \right) E_1 = 0. \quad (8)$$

For the special case of a random or cubic distribution of molecules (in which case $A = 1/3$), Eq. (8) can be simplified by means of the following relationships:

$$(1 + 2p_0)/(1 - p_0) = \epsilon_\infty, \quad (9)$$

$$(1 + 2p_0 + 2p_1)/(1 - p_0 - p_1) = \epsilon_0, \quad (9')$$

$$(\epsilon_\infty + 2)/(\epsilon_0 + 2) = (1 - p_0 - p_1)/(1 - p_0), \quad (10)$$

$$\left(\frac{1 - p_0}{1 - p_0 - p_1} \right) \tau' = \left(\frac{\epsilon_0 + 2}{\epsilon_\infty + 2} \right) \tau' = \tau. \quad (10')$$

Equation (8) then becomes

$$D = D_1 e^{-t/\epsilon_\infty CR} + D_2 e^{-t/\tau} + \epsilon_0 E_1 \quad (11)$$

and introducing the initial conditions, $t = 0$, $D = 0$; $t = 0$, $\frac{dD}{dt} = E_1/CR$, we obtain,

$$D = \epsilon_0 E_1 - \epsilon_\infty E_1 e^{-t/\epsilon_\infty CR} - (\epsilon_0 - \epsilon_\infty) E_1 e^{-t/\tau}. \quad (12)$$

Let q_c be the charge per unit area on either of the condenser plates at any stage of the charging process, that is, at any time t_c after the instant at which the voltage was applied. On converting (12) into e.s.u., we have for the charge at any time during the charging process:

$$q_c = \frac{D_c}{4\pi} = \frac{\epsilon_0 E_1}{4\pi} - \frac{\epsilon_\infty}{4\pi} E_1 e^{-t_c/\epsilon_\infty C_0 R} - \frac{(\epsilon_0 - \epsilon_\infty)}{4\pi} E_1 e^{-t_c/\tau}. \qquad (13)$$

Therefore the ballistic or d-c dielectric constant at any time of charging t_c, is

$$\epsilon(t_c) = \frac{D_c}{E_1} = \epsilon_0 - \epsilon_\infty e^{-t_c/\epsilon_\infty C_0 R} - (\epsilon_0 - \epsilon_\infty) e^{-t_c/\tau}. \qquad (14)$$

The d-c dielectric "constant" appears in this equation as a function of the charging time. Its dependence on charging time is analogous to the dependence of the a-c dielectric "constant" on frequency. The static dielectric constant ϵ_0 is obtained when charging has been continued until $t_c \gg \tau$, that is, when t_c is infinite the dielectric constant has its static value ($t_c = \infty$, $\epsilon(t_c) = \epsilon_0$) and when t_c is zero the dielectric constant is zero ($t_c = 0$, $\epsilon(t_c) = 0$).

The charging *current* is obtained by differentiation of (13) and is

$$\dot{q}_c = \frac{\dot{D}_c}{4\pi} = \frac{E_1}{4\pi C_0 R} e^{-t_c/\epsilon_\infty C_0 R} + \left(\frac{\epsilon_0 - \epsilon_\infty}{4\pi\tau} \right) E_1 e^{-t_c/\tau}. \qquad (15)$$

The charging current per unit voltage gradient, or the apparent conductivity $\gamma_c(t_c)$, is

$$\gamma_c(t_c) = \frac{\dot{q}_c}{E_1} = \gamma_R e^{-t_c/\epsilon_\infty C_0 R} + \gamma_\infty e^{-t_c/\tau}, \qquad (16)$$

where $\gamma_R \equiv (4\pi C_0 R)^{-1}$ and $\gamma_\infty = (\epsilon_0 - \epsilon_\infty)/4\pi\tau$. (It will be noticed that if we define a quantity $G_R \equiv 1/R$, it follows that $(4\pi C_0 R)^{-1} = G_R \cdot d/A = \gamma_R$ since $4\pi C_0 = A/d$. The quantity γ_R is the specific conductance which a fictitious material must possess if it were put in a condenser of geometric capacitance C_0 and required to conduct the same current as C_0 when in series with an external resistance R.)

At any stage of the charging process such that t_c is large as compared with $\epsilon_\infty C_0 R$, but at the same time small as compared with τ the apparent conductivity given by (16) reduces to

$$\gamma_c(t_c) = \gamma_\infty.$$

This value of $\gamma_c(t_c)$, where $\epsilon_\infty C_0 R \ll t_c \ll \tau$, will be designated by γ_0 and called the *initial conductivity*. Using this terminology we see that the *initial conductivity* γ_0 *as determined by d-c measurements equals the infinite-frequency conductivity* γ_∞ *as determined by a-c measurements.*

If t_d represents the time measured from the instant that the voltage is abruptly reduced to zero, and q_d, D_d, $\dot{q}_d$ represent respectively the charge, displacement, and discharge current at time t_d, we have for discharging after *complete* charging

$$q_c(t_d) = \frac{D_d}{4\pi} = \frac{\epsilon_\infty E_1}{4\pi} e^{-t_d/\epsilon_\infty C_0 R} + \frac{(\epsilon_0 - \epsilon_\infty)}{4\pi} E_1 e^{-t_d/\tau} \tag{17}$$

or

$$\epsilon(t_d) = \epsilon_\infty e^{-t_d/\epsilon_\infty C_0 R} + (\epsilon_0 - \epsilon_\infty) e^{-t_d/\tau}. \tag{17a}$$

At the end of the charging process (cf. (13)) or the beginning of the discharge process (cf. (17)) the charge per unit area per unit applied field strength is $\epsilon_0/4\pi$.

The discharge current for *complete* charge is obtained by differentiating (17) with respect to the time:

$$\dot{q}_d = \frac{\dot{D}_d}{4\pi} = -\frac{E_1}{4\pi C_0 R} e^{-t_d/\epsilon_\infty C_0 R} - \frac{(\epsilon_0 - \epsilon_\infty) E_1}{4\pi\tau} e^{-t_d/\tau} \tag{18}$$

or

$$\gamma_d(t_d) = \frac{\dot{D}_d}{4\pi E_1} = \gamma_R e^{-t_d/\epsilon_\infty C_0 R} + \gamma_\infty e^{-t_d/\tau}. \tag{18a}$$

Comparison of (18) and (15), or (18*a*) and (16), shows that for complete charging (that is, t_c effectively infinite or $t_c \gg \tau \gg \epsilon_0 C_0 R$) the charging current vs. time curve is identical, except for direction, with the discharge current vs. time curve. *This is true only of the curves for complete charging and is not true if the polarized condition in the dielectric is not fully formed and the polarization currents are not zero.*

The discharge-current time curve for *incomplete* polarization of the dielectric is not as simple as for complete polarization. When the charging process is broken off before completion, the initial conditions for the discharge are not the same as when charging is complete. The charge at time t_d during a discharge following a charging process which is broken off at t_c is

$$q_d = \frac{D_d}{4\pi} = \frac{\epsilon_\infty E_1}{4\pi} e^{-t_d/\epsilon_\infty C_0 R} - \frac{\epsilon_\infty E_1}{4\pi} e^{-(t_c+t_d)/\epsilon_\infty C_0 R} + \left(\frac{(\epsilon_0 - \epsilon_\infty)}{4\pi} E_1 e^{-t_d/\tau} - \frac{(\epsilon_0 - \epsilon_\infty)}{4\pi} E_1 e^{-(t_c+t_d)/\tau} \right). \tag{19}$$

This is an example of the *superposition principle for residual charges.*

The discharge current for incomplete charge is given by

$$- \dot{q}_d = \frac{-\dot{D}_d}{4\pi} = \frac{E_1}{4\pi C_0 R} e^{-t_d/\epsilon_\infty C_0 R} - \frac{E_1}{4\pi C_0 R} e^{-(t_c+t_d)/\epsilon_\infty C_0 R} + \left(\frac{\epsilon_0 - \epsilon_\infty}{4\pi\tau}\right) E_1 e^{-t_d/\tau} - \left(\frac{\epsilon_0 - \epsilon_\infty}{4\pi\tau}\right) e^{-(t_c+t_d)/\tau}. \quad (20)$$

Or if

$$\frac{-\dot{D}_d}{4\pi} = -\dot{q}_d \equiv I_d$$

$$I_d = \gamma_R E_1 e^{-t_d/\epsilon_\infty C_0 R} - \gamma_R E_1 e^{-(t_c+t_d)/\epsilon_\infty C_0 R} + \gamma_\infty E_1 e^{-t_d/\tau} - \gamma_\infty E_1 e^{-(t_c+t_d)/\tau}. \quad (20a)$$

Equation (20) or (20*a*) is an example of the superposition principle for the residual currents in a dielectric having a single absorptive polarization with a relaxation-time τ. It is evident that in the early stages of the charging process the electronic or instantaneous polarizations responsible for ϵ_∞ have the same external effect as an absorptive polarization because of the fact that the current by which they are formed must flow through the lead resistance R. Thus the condenser acts as though it contained a polarization yielding a dielectric constant with a relaxation-time $\epsilon_\infty C_0 R$.

As the time-constant $\epsilon_\infty C_0 R$ is generally small, the first two terms on the right of (20*a*) may usually be neglected and we have

$$I_d = \gamma_\infty E_1 e^{-t_d/\tau} - \gamma_\infty E_1 e^{-(t_c+t_d)/\tau}. \quad (21)$$

The first term on the right of (21) is the discharge current corresponding to the discharge of the condenser after the residual polarizations have been *fully* formed. The second term gives the value which the charging current would have if the charging process were continued for the interval of time $t_c + t_d$ instead of being discontinued after t_c seconds. When the charging time is large as compared with τ the second term may be neglected and the magnitude of the discharge current is the same function of the discharge time as is the charging current of the charging time; the charging curve and discharge curve can then be superimposed on one another if we disregard the direction of the current. There is only one complete charging current curve and *only one discharging current curve corresponding to complete polarization of the dielectric at any given applied potential. There is, however, an infinite number of discharge curves corresponding to incomplete polarization of the dielectric*, that is, to any time of charging which is shorter than that necessary for complete polarization.

The superposition principle states that any of these discharge curves may be derived from the discharge curve for complete polarization by subtracting from its ordinates the values which the charging current would have if it had continued during the discharge. From the method of deriving equation (21) it is clear that the superposition principle is a necessary consequence of an assumed exponential growth and relaxation of the residual polarizations, as required by the theory of simple anomalous dispersion. If these in fact do not vary exponentially with the time, whatever function they do follow appears in general to obey an empirical superposition rule. Reference to (20*a*) will indicate that if there are m polarizations of different relaxation-times which are quite far apart, each polarization will simply contribute two terms to the expression for I_d; that is,

$$I_d = \sum_{j=1}^{m} (\gamma_{\infty j} E_1 e^{-t_d/\tau_j} - \gamma_{\infty j} e^{-(t_c+t_d)/\tau_j}) + \gamma_R E_1 e^{-t_d/\epsilon_\infty C_0 R} - \gamma_R e^{-(t_c+t_d)/\epsilon_\infty C_0 R}. \quad (22)$$

Thus, the existence of the superposition principle for residual currents as an empirical law suggests that the individual polarizations actually vary exponentially with the time, though direct measurement of the total discharge current seldom gives a single exponential curve. This is, however, not the only possible interpretation.

Abstracts of Technical Articles from Bell System Sources

Radio Telephone System for Harbor and Coastal Services.[1] C. N. ANDERSON and H. M. PRUDEN. Radio telephone service with harbor and coastal vessels is now being given through coastal stations in the vicinities of seven large harbors on the Atlantic and Pacific coasts with additional stations planned. The system is designed to be as simple as possible from both the technical and operating standpoints on both ship and shore.

Recent developments in the shore-station design eliminates all manipulations of the controls by the technical operator. This is made possible principally because of crystal-controlled frequencies on shore and ship, a "vogad" which keeps the transmitting volume of the shore subscriber constant, and a "codan" incorporated in the shore radio receiver which will operate on signal carrier but is highly discriminatory against noise. A signaling system permits the traffic operator to call in an individual boat by dialing the assigned code which rings a bell on the particular boat called. The ship calls the shore station by turning on the transmitter. The radio signal operates the codan in the shore receiver which in turn lights a signal lamp in the traffic switchboard.

Gradually the system has been taking on more and more the aspects of the wire telephone system.

Ship Equipment for Harbor and Coastal Radio Telephone Service.[2] R. S. BAIR. The ultimate objective in the design of radio telephone apparatus for use on ships is to provide equipment which is as convenient and simple to operate as the telephone at home. To a considerable degree this has been accomplished in the new 15- and 50-watt ship sets that have recently been designed for use on harbor craft and coastwise vessels.

The requirements for sets of this type are discussed and the new equipment is described in this paper.

Protective Coatings for Metals.[3] R. M. BURNS and A. E. SCHUH. This book is one of the American Chemical Society Series of Scientific and Technologic Monographs. The chapter headings are: *Protective*

[1] *Proc. I. R. E.*, April 1939.
[2] *Proc. I. R. E.*, April 1939.
[3] Published by Reinhold Publishing Corporation, New York, N. Y., 1939.

Coatings and the Mechanism of Corrosion—Surface Preparation for the Application of Coatings—Types of Metallic Coatings and Methods of Application—Zinc Coating by Hot-Dipping Process—Zinc Coating by Electroplating and Cementation—Protective Value of Zinc Coatings—Cadmium Coatings and their Protective Value—Tin Coatings—Nickel and Chromium Coatings—Coatings of Copper, Lead, Aluminum and Miscellaneous Metals—Coatings of Noble and Rare Metals—Methods of Testing Metallic Coatings—Composition of Paints and Mechanism of Film Formation—The Durability and Evaluation of Paints—Paint Practices—Miscellaneous Coatings.

"The active interest manifested during the past few years in investigations on the general subject of the corrosion of metals has led to the carrying-out of long-time exposure tests which yielded much new basic information having a direct bearing on our knowledge of the useful life of coatings and coated metals. The authors have wisely incorporated a great deal of this information in the discussion of the different types of coatings. Likewise, it has been deemed desirable to devote considerable space to the preliminary preparation of metal surfaces before the application of the coating since the quality of any coating is so dependent upon this factor.

"The new monograph, therefore, covers a much broader field than did the previous one which was really a pioneer in the field of metallic coatings. The investigator of the abstruse problems of corrosion as well as the materials engineer seeking practical help in combating this problem by preventing corrosion by protecting the surface will find this volume a veritable mine of information on all phases of the subject."

A Synthetic Speaker.[4] HOMER DUDLEY, R. R. RIESZ, and S. S. A. WATKINS. This synthetic speaker is an electrical device manipulated by keys and levers for the production of synthetic vocal sounds and their combination into speech. The device was developed as an interesting educational exhibit by the Bell System at the San Francisco Exposition and the New York World's Fair.

From a buzzer-like tone and a hissing noise as raw material, the operator skillfully shapes speech by manipulating the controls to give inflection and the sound spectrum that differentiates one speech sound from another.

This paper covers the development of the device and the training of the operators to demonstrate it.

[4] *Jour. Franklin Institute*, June 1939.

Remotely Controlled Receiver for Radio Telephone Systems.[5] H. B. FISCHER. New radio receiving equipment for shore station used in ship-to-shore telephone circuits has been developed. This equipment is designed to operate on a remotely attended basis and may be located a considerable distance from the telephone terminal equipment. The radio receiver forming a part of the equipment has a codan circuit which operates reliably under high noise conditions and does not require adjustments to compensate for variations in the noise level. An emergency battery power-supply system is provided which is automatically connected to the receiver when the primary alternating-current power supply fails. Power failures are indicated at the telephone central office. A test oscillator which is controlled from the telephone central office is provided which may be used to check the operation of the receiver or to measure the frequency deviations of the incoming signals. The various apparatus units are mounted in two weather-proof cabinets which may be fastened to the same telephone pole which supports the receiving antenna.

Analysis and Measurement of Distortion in Variable-Density Recording.[6] J. G. FRAYNE and R. R. SCOVILLE. Several types of non-linear distortion in variable-density recording are discussed and methods of measurement outlined. The two-frequency inter-modulation method is described. Mathematical and experimental relationships between per cent inter-modulation and per cent harmonic distortion are established. The inter-modulation method is applied to film processing for the determination of optimal negative and positive densities and overall gamma. Variance of these parameters from those indicated by classical sensitometry are traced to halation in the emulsion and to processing irregularities. The use of special anti-halation emulsions appear to reduce residual distortion effects and tend to bridge the gap between inter-modulation and sensitometric control values.

Rubbed Films of Barium Stearate and Stearic Acid.[7] L. H. GERMER and K. H. STORKS. Films of barium stearate and of stearic acid have been prepared on polished chromium and on smooth natural faces of silicon carbide crystals. After these films have been rubbed with clean lens paper, electron diffraction patterns are obtained from them by the reflection method. *Well rubbed films* give patterns characteristic of a single layer of molecules standing with their axes approximately normal

[5] *Proc. I. R. E.*, April 1939.
[6] *Jour. S. M. P. E.*, June 1939.
[7] *Phys. Rev.*, April 1, 1939.

to the surface; the hydrocarbon chains of barium stearate are found to be more precisely oriented than those of stearic acid; exactly the same difference exists between unrubbed single layers of molecules of barium stearate and of stearic acid deposited by the Langmuir-Blodgett method. Thickness of rubbed films on chromium has been found, by the Blodgett optical method, to be the same as that of unrubbed single layers of molecules. *Lightly rubbed films* may be thicker than a single layer of molecules. The arrangement of barium stearate in such thicker films has been found to have been somewhat altered by the rubbing. The axes of the hydrocarbon chains still stand normal to the surface, but lateral arrangement is less regular than it is in unrubbed films of equal thickness. In the case of stearic acid, molecules left on top of the first layer after light rubbing in one direction are found to lie inclined by about 8° to the surface and to point outward against the rubbing direction (Fig. 7); they are arranged in crystals having a structure different from that of the film before rubbing. Such "upset" films of stearic acid are completely removed by very light rubbing in the direction opposite to that of the original rubbing, but they are rather resistant to light rubbing in the same direction.

Diffraction and Refraction of a Horizontally Polarized Electromagnetic Wave over a Spherical Earth.[8] MARION C. GRAY. Formulas are derived for the electromagnetic field at a point on or above the surface of a spherical earth due to the presence of a vertical magnetic dipole. It is shown that the resultant field resembles that due to a vertical electric dipole above a spherical earth of low conductivity, and that in the magnetic case the values of the earth constants are of much less importance than in the electric. Curves are included showing the variation of the field with distance and with height.

Inductive Coordination with Series Sodium Highway Lighting Circuits.[9] H. E. KENT and P. W. BLYE. This paper describes the wave-shape characteristics of the sodium-vapor lamp and discusses the relative inductive influence of various series circuit arrangements in which such lamps are employed. A method is outlined by means of which the noise to be expected in an exposed telephone line may be estimated. Measures are described which may be applied in the telephone plant or in the lighting circuit to assist in the inductive coordination of the two systems. These measures need be considered only when a considerable number of lamps is involved, since noise induction is negli-

[8] *Phil. Mag.*, April 1939.
[9] *Electrical Engineering*, Transactions Section, July 1939.

gible when there are only a few lamps as, for instance, at highway intersections.

Sound Picture Recording and Reproducing Characteristics.[10] D. P. Loye and K. F. Morgan. In the improvement of sound motion pictures, the trend has been to make the response of all parts of the recording and reproducing circuits as nearly "flat" as possible. In some cases, however, this has resulted in unnatural sound, and therefore certain empirical practices have been adopted in the studios and theaters to make pictures sound best.

This paper describes the results of a study the purpose of which has been to evaluate the factors which affect the quality of speech as recorded and reproduced, from the vocal cords of the actor on the sound-stage to the brain of the listener in the theater. The characteristics of the various factors have been determined and combined with dialog, voice effort, and other equalizers designed to produce an overall characteristic " subjectively flat " at the brain of the theater patron. These factors, as well as others which are now in the process of being studied, are presented in this paper.

One of the most important characteristics studied is that of the change in voice quality with a change in the effort on the part of the speaker. This is described in detail in this paper. The stage and set acoustic characteristics, microphone characteristic, and dialog equalization to compensate principally for the hearing characteristic of the average theater listener, are among the factors described herein.

A Dynamic Measurement of the Elastic, Electric and Piezoelectric Constants of Rochelle Salt.[11] W. P. Mason. The elastic, electric and piezoelectric constants of Rochelle salt have been measured at low field strengths by measuring the resonant frequencies and impedance of vibrating crystals. It is shown experimentally that the resonant and anti-resonant frequencies of the crystal are both considerably below the natural mechanical resonant frequency of the crystal in disagreement with the usual derivation of the frequencies of a piezoelectric crystal. By assuming that the piezoelectric stress is proportional to the charge density on the electrodes rather than the potential gradient as usually assumed, theoretical frequencies are obtained which agree with those found experimentally. This theoretical derivation together with the measured frequencies supply values for the piezoelectric constants. The elastic constants measured dynamically show some differences from those measured statically. A large difference is

[10] *Jour. S. M. P. E.*, June 1939.
[11] *Phys. Rev.*, April 15, 1939.

found for the dynamically measured piezoelectric constants from those statically measured, which may be attributed to the finite relaxation time for the piezoelectric elements.

A Vogad for Radio Telephone Circuits.[12] S. B. WRIGHT, S. DOBA, and A. C. DICKIESON. Commercial radio telephone connections must generally be accessible to any telephone in an extensive wire system. Speech signals delivered to the radio terminals for transmission to distant points vary widely in amplitude due to the characteristics of the wire circuits and individual voices. To provide the best margin against atmospheric noise, it is usually the practice to equalize this wide range of speech amplitudes and thus drive the radio telephone transmitter at its full capacity.

Many devices have been proposed to adjust automatically the gain in a circuit to equalize speech volumes. The difficulties of providing a device which will respond properly over a wide range to the complex qualities of a speech signal have only recently been overcome to a satisfactory degree.

The voice-operated gain-adjusting device, or "vogad," described in this paper is a practical design based upon more than a year's experience with one of the most promising devices made available by earlier development effort. A trial installation of this latest vogad is now under way at Norfolk in connection with a new radio telephone system for harbor and coastal service.

[12] *Proc. I. R. E.*, April 1939.

Contributors to this Issue

John R. Carson, B.S., Princeton, 1907; E.E., 1909; M.S., 1912, D.Sc. (Honorary), 1936. American Telephone and Telegraph Company, 1914–34; Bell Telephone Laboratories, 1934–. As Transmission Theory Engineer for the American Telephone and Telegraph Company and later for the Laboratories, Dr. Carson has made substantial contributions to electric circuit and transmission theory and has published extensively on these subjects. The Franklin Institute of Philadelphia recently awarded him the Elliott Cresson Medal. He is now a research mathematician.

J. G. Chaffee, S.B., Massachusetts Institute of Technology, 1923. Western Electric Company, Engineering Department, 1923–25; Bell Telephone Laboratories, 1925–. Mr. Chaffee has been engaged in the study of radio problems at ultra-high frequencies.

W. J. Clarke, B.Chem., Cornell University, 1924; M.A., Columbia University, 1932. Research Laboratory, Devoe and Raynolds Company, 1924–30. Bell Telephone Laboratories, 1930–. Mr. Clarke was at first engaged in studies of organic finishes for telephone equipment, particularly on the compounding of improved finishing materials. More recently his work has been concerned with investigations of molding plastic materials.

Victor E. Legg, B.A., 1920, M.S., 1922, University of Michigan. Research Department, Detroit Edison Company, 1920–21; Bell Telephone Laboratories, 1922–. Mr. Legg has been engaged in the development of magnetic materials and in their applications, particularly for the continuous loading of cables, and for compressed dust cores.

S. O. Morgan, B.S. in Chemistry, Union College, 1922; M.A., Princeton University, 1925; Ph.D., 1928. Western Electric Company, Engineering Department, 1922–24; Bell Telephone Laboratories, 1927–. As Dielectric Research Chemist, Dr. Morgan is concerned with the relation between dielectric properties and chemical composition.

E. J. Murphy, B.S., University of Saskatchewan, Canada, 1918; McGill University, Montreal, 1919–20; Harvard University, 1922–23. Western Electric Company, Engineering Department, 1923–25; Bell

Telephone Laboratories, 1925–. Mr. Murphy's work is largely confined to the study of the electrical properties of dielectrics.

Liss C. Peterson. Chalmers Technical Institute, Gothenburg, 1920; Technical Universities of Charlottenburg and Dresden, 1920–23. American Telephone and Telegraph Company, 1925–30; Bell Telephone Laboratories, 1930–. Mr. Peterson is engaged in amplifier research.

John R. Townsend, Brooklyn Polytechnic Institute. Bethlehem Steel Company, 1915–17. Mathematics and Dynamics Branch, U. S. Ordnance Department, 1917–19. Member of Technical Staff, Western Electric Company, 1919–25; Bell Telephone Laboratories, 1925–. He is now Materials Standards Engineer. He is the author of "Fatigue Studies of Telephone Cable Sheath Alloys," "Physical Properties and Methods of Test for Some Sheet Non-ferrous Metals," and also of many other articles published in technical magazines and discussed before engineering societies. Awarded Dudley Medal, A.S.T.M., 1930.