THE BELL SYSTEM

Technical Journal

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

VOLUME XXXII NOVEMBER 1953 NUMBER 6

THE BELL SYSTEM TECHNICAL JOURNAL

THE BELL SYSTEM TECHNICAL JOURNAL is published six times a year by the American Telephone and Telegraph Company, 195 Broadway, New York 7, N. Y. Cleo F. Craig, President; S. Whitney Landon, Secretary; Alexander L. Stott, Treasurer. Subscriptions are accepted at $3.00 per year. Single copies are 75 cents each. The foreign postage is 65 cents per year or 11 cents per copy. Printed in U. S. A.

THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XXXII NOVEMBER 1953 NUMBER 6

Design Theory of Junction Transistors

By J. M. EARLY

(Manuscript received September 8, 1953)

The small signal ac transmission characteristics of junction transistors are derived from physical structure and bias conditions. Effects of minority carrier flow and of depletion layer capacitances are analyzed for a one dimensional model. The ohmic spreading resistance of the base region of a three dimensional model is then approximated. Short circuit admittances representing minority carrier flow, depletion layer capacitances, and ohmic base resistance elements are then combined into an equivalent circuit. Theoretical calculations are compared to observations for two typical designs.

1.0 INTRODUCTION

1.1 General

Junction transistors have been in commercial production for nearly a year. A detailed understanding of their behavior is necessary both for the increasingly exacting requirements of modern circuit engineering and for the wise design of improved types. Design theory, by relating function to structure, can serve both these needs.

The principal object of this paper is to develop in logical fashion a design theory for junction transistors. The product of the development is an equivalent circuit, founded on device physics, which predicts the circuit characteristics of junction devices in a simple and intelligible fashion. Although attention is concentrated on small signal transmission performance, some large signal aspects are also examined.

1.2 Method and Assumptions

The usefulness of the junction transistor derives primarily from the flow of holes or electrons across two closely-spaced *p-n* junctions, one of which is biased in the forward or conducting direction while the other is biased in the reverse or non-conducting direction. Development of design theory begins quite properly with analysis of this mechanism, which is considered, for simplicity, as a problem in the flow of holes and electrons in one dimension, at right angles to the *p-n* junctions. In the analysis, it is assumed that these carriers are controlled largely by the voltages applied to the junctions and that they move principally by diffusion. The dependence of the diffusion currents on the junction voltages is reduced to a set of two terminal-pair short-circuit admittances, which form the initial and most important segment of the equivalent circuit model for the junction transistor.

Practical transistors have not only the very useful transisting mechanism mentioned above, but also passive capacitances across the charge depletion layers which separate the *p* and *n* regions at each junction. These capacitances limit the useful frequency range of transistors and must be considered in any practical theory. In the synthesis of the equivalent circuit, these capacitances are placed in parallel with the short-circuit input and output admittances which represent the flow of diffusing holes and electrons.

A further limitation on performance is imposed by the ohmic or body spreading resistance of the base region. The base current of the transistor, in flowing from the region between the emitter and collector to the base contact, develops a base contact to emitter voltage which seriously limits the frequency response. Calculation of these effects requires the assumption of flow paths for the base current. The circuit elements representing base spreading resistance effects appear in series in the base leg of the equivalent circuit.

1.3 Existing Design Theory

W. Shockley's classic paper* announcing the junction transistor also initiated the design theory. Diffusion effects for dc and low frequencies were analyzed, and formulae for depletion layer capacitances were developed. The mechanism of the frequency cutoff of the current transmission (alpha) was reported in a subsequent article†, and the effects of

* W. Shockley, The Theory of *p-n* Junctions in Semiconductors and *p-n* Junction Transistors, B.S.T.J., **28**, p. 435.

† W. Shockley, M. Sparks, and G. K. Teal, The *p-n* Junction Transistors, Phys. Rev., **83**, p. 151, July, 1951.

ohmic resistance of the base region were discussed briefly. Still more recently, the dependence of base thickness on collector voltage was used to explain output and feedback effects*. The present paper is both a consolidation and an extension of the earlier works and borrows freely from them. The diffusion current analysis of Appendix A is patterned after Shockley's.

1.4 Scope

The design theory developed here is not complete, even for small signal ac transmission. In particular, effects of large carrier emission densities are not considered, nor are the effects of non-parallel junction arrangements. Despite these omissions, it is hoped that the theory developed will be both useful and instructive to those engineers charged with transistor device and transistor circuit design.

2.0 METHODS AND ASSUMPTIONS

2.1 General

In developing the design theory, it is convenient to break the transistor down into several internal electronic functions and to consider their dependence on structure and materials individually. These functions are then fitted together and used to predict the terminal electrical characteristics. With this approach, it seems proper to describe separately the methods and assumptions used in analyzing each of the functions.

2.2 List of Symbols

The symbols listed here are used in the body of the paper. A separate list for Appendix A appears at the end of that section.

Emitter and collector currents are assumed to flow inward at the corresponding terminals, in accord with the convention usually used for transistors.

a = gradient of $(N_d\text{-}N_a)$, usually given in atoms/cm^4.

a_{ce} = short-circuit forward current transfer constant for theoretical one-dimensional transistor.

C_c = collector to base capacitance with emitter open-circuit ac.

C_{se}, C_{sc} = hole storage or diffusion capacitances at emitter and collector. These capacitances are directly related to the current trans-

* J. M. Early, Effect of Space-Charge Layer Widening in Junction Transistors, I.R.E., Proc., **40**, pp. 1401–1406, Nov., 1952.

mission cutoff frequency and may be used as an alternative characterization of that quantity.

C_{Te}, C_{Tc} = theoretical depletion layer capacitances of emitter and collector.

D, D_p, D_n = diffusion constants for minority carriers, usually given in cm^2/sec.

$f_\alpha = D/\pi w_0^2$ = current transmission or alpha cutoff frequency.

g_{ee}, g_{ce}, g_{ec}, g_{cc} = low-frequency conductance components of y's given below.

h's = set of two terminal-pair parameters, defined by Guillemin, Communication Networks, **2,** p. 137, John Wiley and Sons.

h_{11} = short circuit input impedance.

h_{21} = short circuit forward current transfer ratio.

h_{12} = open circuit feedback voltage ratio.

h_{22} = open circuit output admittance.

I_b = average or dc base current.

I_{pe}, I_{ne}, I_{pc}, I_{nc} = hole and electron components of average or dc emitter and collector currents.

I_{peo} = emitter reverse current when collector is also reverse biased.

J_e = emitter current density in amperes/cm^2.

k = Boltzmann's constant.

kT/q = average thermal energy per carrier, approximately 0.026 electron-volts at 25°C.

L, L_p, L_n = diffusion length or average distance a minority carrier will diffuse before recombining; average distance diffused in one lifetime (τ).

N_A, N_D = concentration of acceptor and donor atoms in semi-conductor, usually in atoms/cm^3.

n = concentration of electrons/cm^3.

n_i = electron concentration which would exist in the semi-conductor at thermal equilibrium if donor and acceptor concentrations were zero.

n_p = thermal equilibrium concentration of electrons in p-region.

p = hole concentration/cm^3.

p_i = hole concentration which would exist in the semi-conductor at thermal equilibrium if donor and acceptor concentrations were zero.

p_n = thermal equilibrium concentration of holes in n-region.

q = electronic charge, 1.6×10^{-19} coulombs.

q/kT = see kT/q.

r_b', r_{b1}', r_{b2}' = ohmic spreading resistances of base region, specifically, the effective base to emitter feedback resistances for diffusion currents and for collector capacitance currents.

r_1, r_2, r_3 = geometrical radii in transistor of Fig. 2(b).

T = temperature in °K.

V_c = average or dc collector to base voltage.

V_c' = electrostatic potential across collector depletion region.

V_s = electrostatic potential across emitter depletion layer at therma equilibrium (no biases applied).

v_c = small signal ac collector to base voltage.

w, w_0 = base region thickness.

w_1, w_2, w_3 = base region thicknesses in transistor of Fig. 2(b).

x_m = thickness of collector depletion region.

y_{ce}, y_{ce}, y_{ec}, y_{cc} = theoretical short circuit input, forward transfer, feedback, and output admittances for one-dimensional transistor.

α, α_0 = short-circuit emitter to collector current transfer ratio and its low-frequency value.

α^*, α_0^* = collector junction current multiplication ratio and its low-frequency value.

β, β_0 = current transport ratio across base region and its low-frequency value.

γ, γ_0 = current emission ratio at emitter and its low-frequency value.

ϵ_0 = dielectric constant of vacuum, 8.854×10^{-14} farad/cm.

k = relative dielectric constant, ϵ/ϵ_0.

ρ, ρ_b = resistivity, base region resistivity.

σ_{nc}, σ_{pc} = conductivities produced by electrons and holes in collector region.

τ, τ_n, τ_p = lifetimes of minority carriers.

μ_{bc} = constant of feedback generator used to characterize modulation of dc base spreading resistance.

$\omega = 2\pi f$ = angular frequency in radians.

$\omega_\alpha = 2\pi f_\alpha = 2D/w_0^2$ = alpha or current transmission cutoff frequency in radians.

2.3 Minority Carrier Admittances

An admittance representation of minority carrier diffusion is a way of writing the dependence of the diffusion currents on the junction potentials. To obtain this dependence analytically, the minority carrier densities on both sides of each of the two depletion layers (emitter and collector) are assumed to be exponential functions of the junction voltages. This exponential dependence is a result of the normal thermal distribution of hole and electron energies. The carrier diffusion currents are computed directly from the gradients of the minority carrier densities at the depletion layer surfaces. Since the gradients of the carrier densities are affected by many conditions besides the junction voltages, additional assumptions are necessary. Their nature and pertinence may be seen from consideration of the normal operation of a junction transistor.

The three principal regions of a junction transistor, the emitter, the base or control, and the collector, are indicated in Fig. 1. These regions are separated by transition regions in which the conductivity type changes either gradually or abruptly from p-type to n-type. Roughly coincident with these transition regions are the emitter and collector depletion layers across which the emitter and collector voltages appear

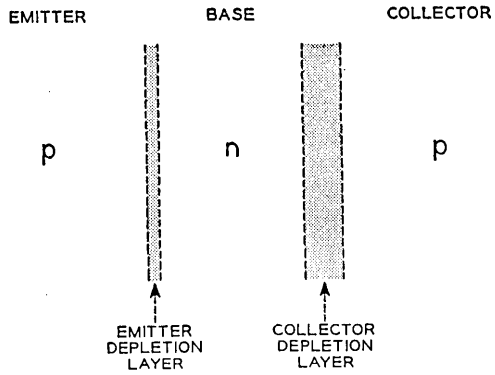


Fig. 1 — p-n-p transistor.

when the unit is biased. In normal operation for the p-n-p transistor shown, the emitter is biased positive with respect to the base so that a current of holes is injected into the base from the emitter. The collector is biased negative with respect to the base so that the holes diffusing across the base from the emitter are collected whenever they reach the edge of the collector depletion layer.

In the analysis each of the three major regions is assumed to have a uniform resistivity, ρ; a diffusion constant for minority carriers, D, which is a measure of the speed with which injected carriers will diffuse; and a lifetime for minority carriers, τ. This lifetime is the average time which a minority carrier remains free before recombining with a majority carrier. The minority carrier density in each region is assumed to have a thermal equilibrium value in the absence of applied potentials. The density is increased or decreased exponentially from this value by the applied potentials. The base layer is assumed to have a thickness, w, which is dependent on the collector voltage V_c. An increase of collector voltage increases the collector depletion region thickness, x_m, thus decreasing the base thickness. The rate at which base thickness changes with collector voltage is determined by the nature of the transition from base to collector. For gradual transitions, the rate of the transition is important, while for abrupt or step transitions the rate of change of base thickness with collector voltage is determined by the base region and collector region resistivities.

Determination of the diffusion currents from minority carrier density gradients requires determination of minority carrier densities everywhere in the three principal regions of Fig. 1. These are obtained by solving a continuity equation for carrier flow in each region, subject to the applied junction potentials and other assumptions described above. It must be pointed out that, in normal operation, there may be a significant flow of electrons to the emitter and from the collector in the p-n-p transistor of Fig. 1. Small signal ac diffusion currents are determined by assuming small signal variations of the junction voltages and discarding all but first-order ac terms from the diffusion currents.

Results of the analysis are given in Section 3.0, and the analysis appears in Appendix A.

2.4 Depletion Layer Capacitances

In a p-n junction with no bias potential applied, there is a tendency for holes to diffuse into the n-region and for electrons to diffuse into the p-region. This creates a slight unbalance of charge in the two regions

and the resulting electrostatic potential keeps each type of carrier in its own region. The potential appears across a thin layer separating the two regions. In this depletion layer, the hole density is lower than in the p-region and the electron density is lower than in the n-region, and there is a net charge density. Acceptor and donor atoms are not neutralized by mobile charge as they are in the p- and n-regions, but instead serve to terminate the field of the electrostatic potential. Application of external potential across the junction changes the electrostatic potential, and by exposing more or fewer fixed (donor and acceptor in equal number) charges widens or narrows the depletion layer.

The passive capacitance of this region is simply that of a parallel plate condenser having a plate spacing equal to the layer thickness. Calculation of this capacitance is explained in Section 3.0, following the discussion of the minority carrier diffusion admittances.

*2.5 Base Spreading Resistance**

Implicit in the one-dimensional analyses described above is the assumption that the base region is everywhere at the same potential. Actually, since the emitter and collector currents are not equal, current must flow through the base region parallel to the junctions. Because the base region has finite, rather than zero resistivity, this current produces transverse voltage drop in the base region.

It is assumed that the most important effect of these voltage drops is the feedback produced between the base contact and the emitter junction. In consequence, each of the ohmic base resistances studied is defined as the quotient of an average voltage between base contact and emitter junction divided by the current producing it. The need for defining more than one feedback base spreading resistance results from the fact that the base current has two principal ac components. One of these is the difference between the emitter and collector minority carrier diffusion currents. The other is the collector depletion layer capacitance current. The feedback effects of these two currents on the emitter junction are the same only when the flow paths of the two currents through the base region are the same. Consequently, the representation of base resistance effects is somewhat more complicated in transistors where the flow paths differ than in those where they are identical or nearly so.

* The majority carrier resistance of the base region for base current flow parallel to the junctions. The word "spreading" was suggested by the base contact geometry of Fig. 2(a) and readily distinguishes this resistance from the "base resistance" of the familiar tee network, which was long believed to be identical with it. It is not. See Sections 5.2 and 5.3.

Fig. 2(a) shows a structure for which the flow paths to the base contact are substantially the same for all components of the base current. Both the collector capacitance current and the diffusion loss base current enter the base region substantially uniformly over the entire area and follow the same path to the base contact. In Fig. 2(b) these two currents have quite different flowpaths and the associated feedback resistances are likewise very different. The general method of calculation is, however, the same in both cases.

Another important effect is associated with modulation of the dc voltage drop in the base region. The base current ordinarily has a dc as well as an ac component, and a dc voltage drop occurs between the base contact and the emitter junction. Since the base region thickness changes when collector voltage changes, the dc resistance of the base region is modulated by the collector voltage, producing a modulation of the dc voltage between base contact and emitter.* This effect is most easily represented by an ac voltage generator in series with the base

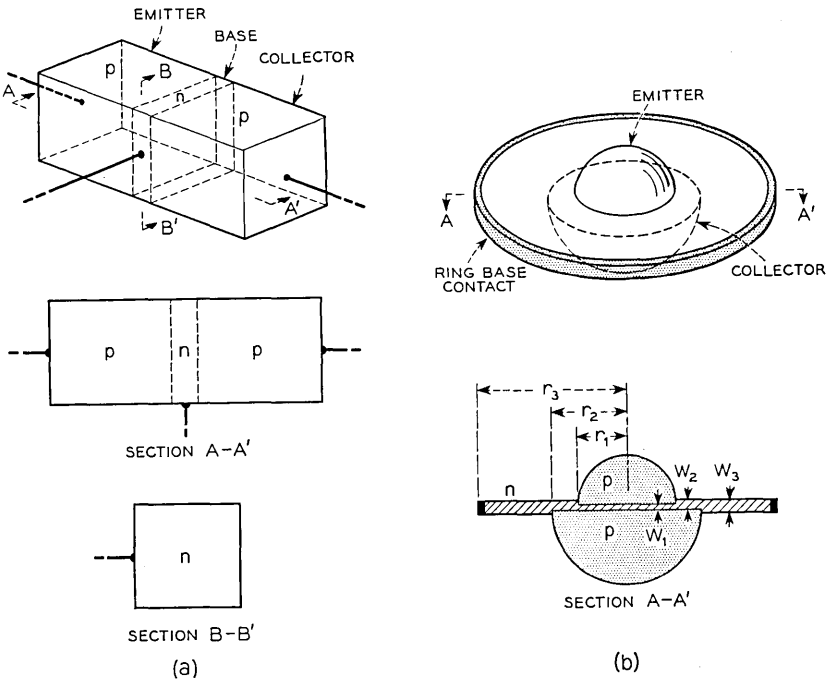


Fig. 2 — *p-n-p* transistor structures.

* This effect was first pointed out by J. N. Shive.

contact. The voltage is computed as the product of the dc base current and the modulation of the dc base resistance. This modulation can be calculated from the base region resistivity and the dependence of base thickness on collector potential.

2.5 Summary of Methods

In developing the design theory, simple physical assumptions are made concerning the behavior of the charge carriers in the semiconductor. The transistor is studied as a one-dimensional problem and the per unit area electrical characteristics of the one dimensional structure are computed. The effects of current flow within the base region parallel to the junctions are then calculated for a three-dimensional model. Finally, the equivalent circuit representations of these electronic functions are combined in structural fashion to give the terminal electrical characteristics of the junction transistor triode.

It should be noted that the base region thickness between emitter and collector is assumed uniform, and that design theory has not been extended here to cover the case of non-uniform thickness. Likewise, edge effects at the emitter and surface effects in general are neglected. These omissions were made for mathematical simplicity and are necessary omissions in a one-dimensional analysis. The place of surface leakage among the electronic functions is discussed at the end of Section 4.0. Analysis of the effects of sharp discontinuities in base layer thickness requires new solutions to the continuity equation but gradual changes in thickness can be accounted for by averaging over the active area of the transistor the short circuit admittances which are the subject of the next section.

3.0 ONE-DIMENSIONAL TRANSISTOR

3.1 General

This section deals with the small signal transmission electronics of the structure of Fig. 1. It is assumed that the emitter is biased to provide a flow of carriers into the base and that the collector is reverse biased sufficiently so that no majority carriers can diffuse out of the collector region into the base region (a reverse voltage of 0.5 volts is more than enough to prevent this). The four admittances associated with minority carrier flow and the two depletion layer capacitances are indicated in Fig. 3. In each case, the design expressions are given first in their most exact form and are progressively simplified. For convenience in discussion and comparison, current densities per unit area rather than

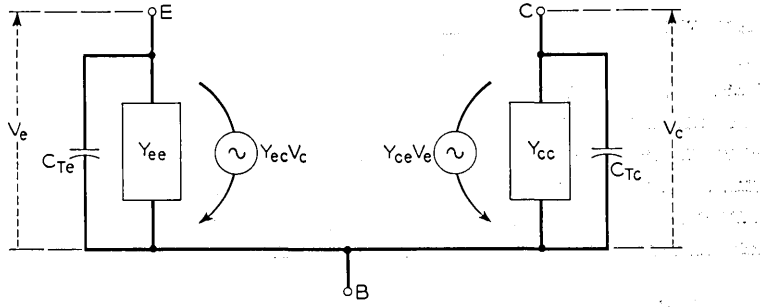


Fig. 3 — Theoretical equivalent circuit for "one-dimensional" transistor.

currents are used and the admittances and barrier capacitances are written on a per square centimeter basis. It will be noted that transverse voltage drops resulting from base spreading resistance are ignored in developing the expressions of this section.

A number of physical mechanisms are involved in the admittances for minority carrier flow. The forward current across the emitter junction rises as an exponential of the emitter to base region voltage. This is the result of the thermal energy distribution of minority carriers and is common to all thermionic emission. A natural effect of this exponential dependence is that a given change in the voltage results in a fixed per cent change in the current, thus producing an ac admittance which is proportional to the average or dc emitter current. For several reasons not all of the current which flows through the emitter junction is collected. First, some of the emitter current consists of electrons diffusing into the emitter body and of displacement current through the emitter depletion layer capacitance. These effects are expressed in the emission factor or γ, which is the ratio of the hole current injected into the base region to the total emitter current. Further, some of the injected holes recombine in the base layer. Those which are collected suffer a transit delay which results in a phase difference between emitter and collector currents. These two effects are summed up in the forward current transport factor or β, which is the ratio of the minority carrier current reaching the collector to that injected by the emitter. Finally, the current of holes entering the collector from the base gives rise to a much smaller flow of electrons from the collector to the base, thus producing a collector multiplication factor or α^*, which is the ratio of total carrier current crossing the collector junction to the hole current entering it from the base.

Although the most important minority carrier flow originates at the emitter, this flow is altered by changes in collector reverse voltage. As the collector reverse potential is increased, the base region becomes

narrower because of widening of the collector depletion layer. This reduction in base thickness permits more emitter current to flow for a fixed emitter to base voltage, so that both emitter and collector currents are increased, thus producing output and feedback admittances. However, since large changes in collector potential are required to produce small changes in base thickness, relatively small changes in the junction currents are produced and the admittances are far smaller than those associated with change of emitter potential.

3.2 Diffusion Current Admittances

The short circuit two terminal-pair admittances associated with the diffusion of minority carriers in the structure of Fig. 1 are

$$y_{ee} = \frac{q}{kT}\left[(I_{pe} - I_{peo})\frac{(1 + i\omega\tau_p)^{1/2}\tanh w_0/(D_p\tau_p)^{1/2}}{\tanh\left[(1 + i\omega\tau_p)^{1/2}w_0/(D_p\tau_p)^{1/2}\right]} + I_{ne}(1 + i\omega\tau_{ne})^{1/2}\right]$$

$$y_{ce} = -\frac{q}{kT}\left[(I_{pe} - I_{peo})\left[\frac{(1 + i\omega\tau_p)^{1/2}\tanh w_0/(D_p\tau_p)^{1/2}}{\sinh\left[(1 + i\omega\tau_p)^{1/2}w_0/(D_p\tau_p)^{1/2}\right]}\right]\left[1 + \frac{\sigma_{nc}}{\sigma_{pc}}\right]\right.$$

$$y_{ec} = -\frac{\partial w}{\partial V_c}\frac{(1 + i\omega\tau_p)^{1/2}}{(D_p\tau_p)^{1/2}\sinh\left[(1 + i\omega\tau_p)^{1/2}w_0/(D_p\tau_p)^{1/2}\right]}I_{pc}$$

$$y_{cc} = -\frac{\partial w}{\partial V_c}\frac{(1 + i\omega\tau_p)^{1/2}}{(D_p\tau_p)^{1/2}\tanh\left[(1 + i\omega\tau_p)^{1/2}w_0/(D_p\tau_p)^{1/2}\right]}I_{pc}\left[1 + \frac{\sigma_{nc}}{\sigma_{pc}}\right]$$

in which I_{pe}, I_{ne}, and I_{pc} are average hole and electron currents at the emitter and collector junctions and I_{pe0} is the emitter reverse current measured with both junctions reverse biased. w_0 is the (time) average base region thickness and V_c is the average collector voltage. The depletion layer capacitances which shunt the input and output diffusion admittances, y_{ee} and y_{cc}, are discussed in Section 3.3.

Some general features of the admittances may be noted at once. The expressions are similar to those for a section of lossy transmission line, but differ significantly in the ordering of the magnitudes of the terms. Each of the admittances is proportional to a dc current, and in fact, since I_{pc} is ordinarily ninety per cent or more of I_{pe}, approximately the same dc current. This effect, which results from the exponential dependence of emitter current on emitter potential, can be related to the transmission line analogy by the argument that the increase in minority carrier density in the base region which accompanies increase of dc current lowers the characteristic impedance of the transmission line.

Finally, the last two terms, the feedback and output admittances, are smaller than the first two by approximately

$$\frac{kT}{qw_0}\frac{\partial w}{\partial V_c}$$

which is usually 10^{-3} or less, but the four form a rather symmetrical set.

The symmetry of the terms can be seen by removing the dissymmetries. The term $I_{ne}\ (1 + i\omega\tau_{ne})$ in y_{ee} results from diffusion of electrons into the emitter from the base and is the emission loss term which makes γ less than unity. The factor $(1 + \sigma_{nc}/\sigma_{pc})$ which appears in y_{ce} and y_{cc} is the collector multiplication factor α^* which results from the flow of electrons out of the collector body. If γ and α^* are assumed to be unity, the admittances assume a more symmetrical form. In that case $y_{ce}/y_{ee} = y_{ec}/y_{cc} = -\beta$, the base transport factor, and the forward and reverse current transmission ratios are identical. This can be seen more clearly if the hyperbolic expressions are replaced by the first two terms of their polynomial expansions, yielding

$$y_{ee} = g_{ee}\frac{(1 + i\omega/\omega_\alpha)}{(1 + i\omega/3\omega_\alpha)}$$

$$y_{ce} = \frac{-\beta_0 g_{ee}}{1 + i\omega/3\omega_\alpha}$$

$$y_{ec} = \frac{-\beta_0 g_{cc}}{1 + i\omega/3\omega_\alpha}$$

$$y_{cc} = g_{cc}\frac{(1 + i\omega/\omega_\alpha)}{1 + i\omega/3\omega_\alpha}$$

where

$$g_{ee} = qI_e/kT, \qquad g_{cc} = -\frac{1}{w}\frac{\partial w}{\partial V_c} I_{pc}\left(1 + \frac{1}{2}\frac{w_0^2}{D_p\tau_p}\right),$$

$$\omega_\alpha = \frac{2D_p}{w_0^2} \quad \text{and} \quad \beta_0 = 1 - \frac{1}{2}\frac{w_0^2}{D_p\tau_p}$$

is the low frequency value of the base transport factor. A lumped parameter equivalent circuit for these simplified admittances is shown in detail in Fig. 4(a).

It is apparent here that a single parameter, $\omega_\alpha = 2D_p/w_0^2$, specifies the frequency variation of all four admittances. This frequency dependence, which has been commonly measured as the alpha cutoff frequency

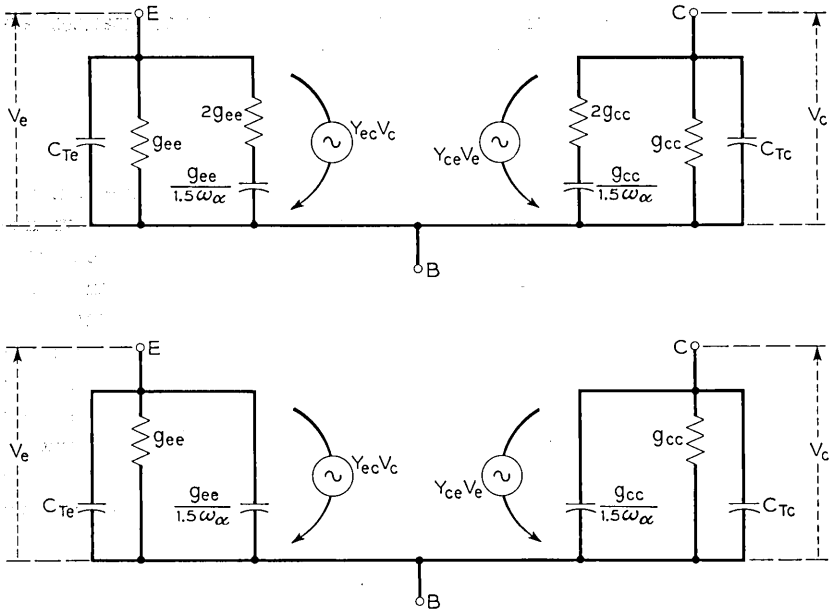


Fig. 4 — Simplified equivalent circuits for "one-dimensional" transistor.

or three db point† of the current transmission amplitude, appears in the admittances in the form of storage capacitance or as transfer delay. These result from minority carrier storage in the base region and are of the same nature as hole storage capacitances in semi-conductor diodes.

If desired, further simplication can be obtained by eliminating the effective power loss in the storage capacitances as shown in Fig. 4(b). In this arrangement, as in the more rigorous forms preceding it, the forward transfer admittance y_{ce} is a little smaller than the input admittance y_{ee} and the feedback admittance y_{ec} is not quite as large as the output admittance y_{cc}. This makes the circuit delta $(y_{ee}y_{cc} - y_{ce}y_{ec})$ greater than zero so that the structure is inherently stable when γ and α^* are unity.

3.3 Depletion Layer Capacitances

Even in the absence of applied potentials, there exists at useful *p-n* junctions a depletion layer in which the density of mobile carriers is low,

† R. L. Pritchard has pointed out that this three db point is 22 per cent larger than ω_α as defined here. See R. L. Pritchard, Frequency Variations of Current-Amplification Factor for Junction Transistors, Proc. I.R.E., **40**, p. 1476, Nov., 1952.

and across which there is a small electrostatic potential. The potential difference across the region exists primarily because of the difference in energy levels between the conduction band in which the mobile electrons exist and the valence bond band in which the mobile holes move. The potential difference depends on the densities of holes and electrons in the p and n regions, but it cannot exceed the energy gap or difference in band levels, so long as the junction is at equilibrium.

If a reverse voltage is applied to a junction, the applied voltage appears principally across the depletion region, where it strengthens the electric field by widening the barrier region so as to bring more fixed donor and acceptor charges into the field. It is obvious that the depletion region has a capacitance, since an electric field exists across it. This capacitance decreases with increase of reverse voltage, since the capacitance is charged not by bringing mobile carriers to fixed electrodes but rather by widening the region to include new fixed charges from the semiconductor regions on each side.

Both emitter and collector capacitances may be calculated easily for the principal cases of published engineering interest, the graded transition and the abrupt or step transition. For graded transitions, the depletion layer is in a region of linearly changing fixed charge density (zero at the center of the layer). The step junction has the barrier layer almost entirely in either a p or an n region of uniform fixed charged density since the fixed charged density in the other region is usually so large that the field effectively terminates at its surface.

The general expression for barrier capacitance is

$$C = \frac{\kappa\epsilon_0}{x_m}$$

where x_m is barrier thickness and the other symbols have conventional meanings. In the case of graded junctions, this becomes

$$C = \frac{\kappa\epsilon_0}{2}\left(\frac{2qa}{3\kappa\epsilon_0 V_c'}\right)^{1/3}$$

where a is the rate of change of fixed charge density in charges per cm^3 per cm. For step junctions, the relation is

$$C = \kappa\epsilon_0\left(\frac{q(N_d - N_a)}{2\kappa\epsilon_0 V_c'}\right)^{1/2}$$

where $(N_d\text{-}N_a)$ is the fixed charge density in the low charge density region, ordinarily the base region in transistors. The potential V_c is the electrostatic potential across the depletion layer and under bias conditions is

the sum of the equilibrium barrier potential and whatever part of an applied external potential is not lost in IR drops in the p and n regions. For reverse bias conditions, the sum of the equilibrium potential and the entire applied potential is almost always a satisfactory approximation, but for equilibrium and forward bias conditions, the barrier potential may be computed more easily from the minority carrier densities immediately adjacent to the depletion layer.

The equilibrium barrier potential may be calculated by

$$V_s \simeq \frac{kT}{q} \ell n \frac{n_i}{n_p} \cdot \frac{p_i}{p_n}$$

in which n_n, p_p and n_i, p_i are, respectively, equilibrium and intrinsic carrier densities in the two regions. The total depletion region potential for a forward biased emitter junction may be calculated by the same expression but the hole density in the base region side of the emitter depletion layer is not the equilibrium value p_n but rather

$$p = \frac{w_0 J_e}{qD_p}$$

which is simply a modified form of the equation for diffusion current. Because of the large difference in the voltages across the barriers, emitter depletion region capacitance per unit area is ordinarily 2–20 times larger than the collector depletion region capacitance per unit area. It should not be assumed that emitter depletion region capacitance is the more important. It contributes but a small fraction of the emitter admittance, while the collector depletion layer capacitance contributes greatly to collector admittance.

3.4 Summary

A one-dimensional study of the small signal transmission properties of the junction transistor shows two terminal-pair short circuit admittances which are closely proportional to the dc bias currents. The input and output admittances of this set are shunted by depletion layer capacitances which are essentially passive circuit elements. A major feature of the diffusion current admittances is the presence of a single frequency determining factor, the alpha cutoff frequency. Comparison of the diffusion or storage capacitances with the depletion region capacitances shows that the collector depletion region capacitance is usually much larger than its storage capacitance, while at the emitter the reverse is true. This results from the fact that the output and feedback minority

carrier admittances are many orders of magnitude smaller than the input and forward transfer terms. The output and feedback admittances decrease with increase of collector reverse voltage in much the same way that collector capacitance decreases, while the input and forward transfer admittances are but little affected.

This simple picture of the junction transistor is incomplete in that it ignores the important effects of majority carrier resistance in the base region. These effects are discussed in the next section.

4.0 EFFECTS OF BASE REGION RESISTANCE

4.1 General

The resistance of the base region to the flow of currents parallel to the emitter and collector junctions is important primarily because of voltage developed between the base contact and the emitter junction. Necessarily associated with this is power dissipation which is usually, however, of less importance than the feedback effect of the voltage.

The primary factors in determination of base region feedback voltages are the materials and geometry of the base region and the flow paths for the transverse currents moving through the base region to the base contact. Reduction of the base region resistance effects to equivalent circuit elements permits them to be incorporated in the equivalent circuit obtained in the previous section. This circuit is then an essentially complete model for the electronic mechanisms or functions in junction transistor triodes.

4.2 Base Region Currents

Two of the three principal components of the base current have nearly identical origin and flow paths, while the third component differs in origin and may differ greatly in flow paths to the base contact. The dc component of the base current arises principally from recombination of injected holes in the portion of the base between emitter and collector,* as does also the ac component associated with the diffusion admittances. The ac component of base current required to charge the collector barrier capacitance, however, is introduced into the base uniformly over the surface of the collector. The ac component required to charge the emitter capacitance is both small and similar in flow paths to the first components discussed.

* In some transistors, much of the recombination occurs on the exposed surface of the base region. This surface recombination may be replaced by a reduction of volume lifetime for a one-dimensional analysis. This is not exact, but is a fair approximation.

For the structure of Fig. 2(a), all current components are introduced into the base layer relatively uniformly over its area. The dc component and the $(1 - \alpha)$ or diffusion component originate within the layer volume, while the collector capacitance current is introduced from the collector side of the layer, but this difference is of little consequence. The transverse current density is largest at the base contact and diminishes monotonically to zero at the layer edge opposite the contact. The feedback voltage to the emitter junction consequently rises most rapidly close to the contact and becomes nearly constant as the opposite side of the base region is reached.

For the structure of Fig. 2(b), significant differences in flow paths are apparent. The transverse current densities are zero at the center of the unit and rise out to the edge of the emitter. Here the dc and $(1 - \alpha)$ components of current density begin to decrease because of the increasing cross-section of the base region, while the collector capacitance components continues to increase because of the additional contribution from the larger collector area. All components decrease in density beyond the collector circumference. It is apparent that two feedback resistances are required to describe the separate electronic functions in this case.

4.3 Effective Feedback Resistances

The resistance effects of importance are not the series resistances to the flow of the various current components through the base region to the base contact, but rather the feedback resistances defined by the quotient of the average base contact to emitter junction voltage by the current component producing the voltage. The calculations necessary to determine such resistances are detailed in Appendix II.

For the structure of Fig. 2(a), a single equivalent resistance given by

$$r_b' = f\left(\frac{\rho_b}{w}\right) \simeq \frac{\rho_b}{w}$$

is sufficient. Both $(1 - \alpha)$ and C_c current components pass through this resistance en route to the base terminal. In the structure of Fig. 2(b), the $(1 - \alpha)$ or diffusion current component has an effective feedback resistance of†

$$r_{b_1}' \simeq \rho_b \left[\frac{1}{8\pi w_1} + \frac{1}{2\pi w_2} \ell n(r_2/r_1) + \frac{1}{2\pi w_3} \ell n(r_3/r_2)\right]$$

while the collector capacitance component of base current has associated

† This value of r_{b1}' is an upper bound, since the assumptions are pessimistic.

with it a resistance

$$r'_{b_2} = \rho_b \left(\frac{1}{2\pi w_3} \ell n(r_3/r_2) + (r_\perp/r_2)^2 \left[\frac{1}{2\pi w_2} \ell n \frac{r_2}{r_1} + \frac{1}{8\pi} w_1 \right] + \left[1 - \left(\frac{r_1}{r_2} \right)^2 \right] \frac{1}{4\pi w_2} \right)$$

For the proportions shown, the ratio of r'_{b_1} to r'_{b_2} is approximately 1.5 to one. Reasonable variations in the proportions can make this ratio as small as unity or as large as five.

A common feature in these resistances is the presence of the base region thickness w in the denominator. Since low values of base spreading resistance are desirable, this requirement opposes directly the primary requirement for a high alpha cutoff frequency — a thin base layer.

4.4 Base Resistance Modulation Feedback

As was mentioned previously, the dc voltage drop between the base contact and the emitter junction is modulated by the widening and narrowing of the collector depletion layer which alternately increases and decreases the dc base resistance. The resulting ac voltage may be represented by a voltage generator in series with the base resistance, as

$$\mu_{bc} v_c = I_b \frac{\partial r'_b}{\partial w} \frac{\partial w}{\partial V_c} v_c$$

where I_b is the dc base current, V_c is the dc collector voltage, v_c is the ac collector voltage, and $\mu_{bc} v_c$ is the feedback voltage.

For the structure of Fig. 2(a), the expression given earlier for r_b' may be differentiated to obtain the $\partial r_b'/\partial w$. The calculation of $\partial w/\partial V_c$ has been indicated in Section 3.0. The resulting expression for μ_{bc} is approximately

$$\mu_{bc} \simeq I_b \frac{r'_b}{w} \frac{\partial w}{\partial V_c}$$

Determination of this feedback effect for the structure of Fig. 2(b) is more difficult. Only those portions of the base region adjacent to the collector barrier are affected by the barrier widening, so that only the first two terms in the expression for r_b', given above enter into the calculation. The value of μ_{bc} is

$$\mu_{bc} \simeq I_b \rho_b \left(\frac{1}{8\pi w_1^2} + \frac{1}{2\pi w_2^2} \ell n \frac{r_2}{r_1} \right) \frac{\partial w}{\partial V_c}$$

This feedback term cannot be reduced to an expression containing r_b',

in the same simple way possible for the structure of Fig. 2(a), since the base resistance between the collector circumference and the base contact is not changed by collector depletion region thickness changes.

The details of the base contact may change the magnitude of this feedback greatly. In particular, poor placement of a bonded base contact may sometimes result in very large values of μ_{bc}. A point of general interest is that the sign of the feedback is determined directly by the direction of base current flow, while the phase is otherwise exactly that of the collector voltage.

4.5 Equivalent Circuit with Base Resistance Effects

The base resistance effects discussed above can be combined with the admittance circuits obtained in the one-dimensional study of Section 3.0 to give equivalent circuits for the three-dimensional structures of Fig. 2. The resulting representations, shown in Fig. 5, reflect the geometry of these structures since the base resistance elements are shown in a branch through which the base current must flow to reach the base terminal. All of the effects can be seen to give feedback, which is to say, coupling from the output circuit of the collector to the input loop of the emitter. This leads to complication of the electrical characteristics of the transistor, whose characteristics would otherwise be given by the electronic functions developed in the previous section.

Fig. 5(a) is an equivalent circuit which represents the small signal transmission properties of the structure of Fig. 2(a). It should be noted that a single base resistance is needed, through which all of the base current passes. The effects of surface leakage are indicated by the admittance Y_l from collector terminal to base resistance. The indicated uncertainty of placement of this leakage effect with respect to r_b' reflects the fact that the feedback resulting from the leakage depends on the position of the leakage with respect to the base contact.

The structure of Fig. 2(b) is represented by the circuit of Fig. 5(b). The separate base resistances r_{b_1}' and r_{b_2}' mirror the physical fact that the collector region is, on the average, closer to the base contact than is the emitter region. Leakage effects may be added to this circuit in the same manner as before.

Although the circuits of Fig. 5 represent primarily small signal ac transmission functions in the transistor, very similar representations may be employed for large signal functions. The principal changes occur in the diffusion admittances, which should be replaced by the diffusion currents written as functions of the barrier voltages, and in the capaci-

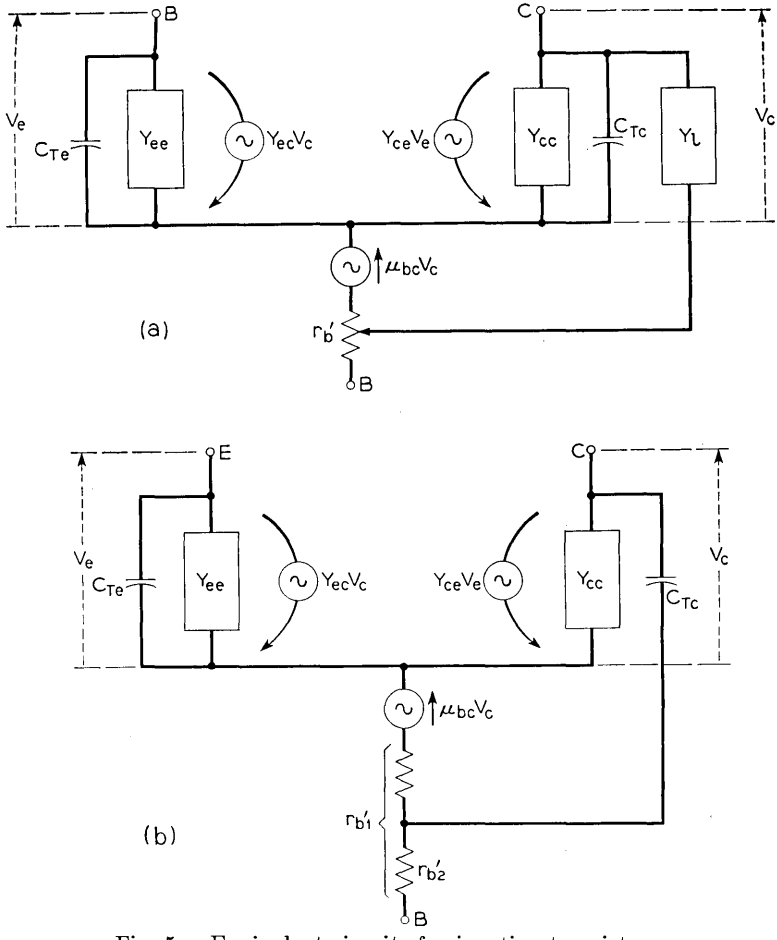


Fig. 5 — Equivalent circuits for junction transistors.

tances, whose variation with barrier potential must now be considered explicitly.

5.0 DESIGN CALCULATIONS

5.1 General

The understanding of junction transistors is now sufficiently advanced that this type of device is designable — the engineer may work from requirements to choice of structure and material. The limitations on designability are now primarily lack of adequate control over manufacturing variables together with the inherent characteristics of the

semi-conductor materials available to the designer. The designability of the major small signal transmission parameters of junction devices is illustrated below by a comparison of calculated and measured characteristics of both grown and fused junction devices.

The characteristics calculated are the theoretical short circuit conductances and current transmission ratios (alphas), the current transmission three db cutoff frequencies, the collector capacitances, the base region spreading resistances, and the low-frequency values of the "h" parameters. The latter are computed because of their ease of measurement and interpretation and are compared with measured values.

5.2 n-p-n Drawn Junction Transistor

The properties assigned to this unit are believed to be reasonable, but a close check on many of them is very difficult. Base region thickness was chosen as a reasonable value for current practice. The physical structure is that of Fig. 2(a), with the bar cross-section a square 20 mils on a side. It should be noted that the assumptions made are consistent; e.g., the grading of the collector junction is consistent with the collector and base region resistivities.

The physical parameters of interest are:

V_c Collector voltage, 4.5 volts
I_e Emitter current, -1 ma
w Base thickness (not including depletion layers), 3.6×10^{-3} cm or 1.41 mil
ρ_b Base resistivity, 1 ohm-cm
τ_b Electron lifetime in base, 10 μ sec
ρ_e Emitter resistivity, 0.01 ohm-cm
τ_e Hole lifetime in emitter, 0.45 μ sec
ρ_c Collector resistivity, 1.7 ohm-cm
τ_c Hole lifetime in collector, 30 μ sec
a Concentration gradient in junctions, 3×10^{18} atoms/cm^4
A Junction areas, 0.0025 cm^2

The parameters of this unit are then

$$g_{ee} = \frac{q}{kT} I_e = \frac{10^{-3}}{0.026} = 0.039 \text{ mhos}$$

$$\gamma_0 \simeq \frac{2}{1 + \frac{\rho_e}{\rho_b} w/\sqrt{D_p \tau_e}} = \frac{1}{1 + 0.008} = 0.992$$

$$\beta_0 \simeq \left[1 - \frac{1}{2} \frac{w^2}{D_n \tau_b}\right] = 1 - 0.0072 \simeq 0.993$$

$$\text{at } 25°\text{C } \alpha_0^* \simeq \left[1 + \frac{1}{2}\frac{\sigma_{pc}}{\sigma_{nc}}\right] = 1 + 0.0003 \simeq 1.0003$$

$$70°\text{C } \alpha_0^* \simeq 1.03$$

$$25°\text{C } a_{ce} = \gamma_0\beta_0\alpha_0^* = 0.992 \times 0.993 \times 1.0003 = 0.985$$

$$70°\text{C } a_{ce} = 1.015$$

$$\text{Now } \frac{\partial w}{\partial V_c} = \frac{x_m}{6V_c} = \frac{6 \times 10^{-4}}{6 \times 6} = 1.6 \times 10^{-5} \text{ cm/volt}$$

$$g_{cc} = -\frac{1}{w}\frac{\partial w}{V_c\partial} I_c \cosh (w/L_n)$$

$$\simeq \frac{kT}{qw}\frac{\partial w}{\partial V_c} g_{ee'} = 0.026 \times 1.6 \times 10^{-5} \times 0.039$$

$$= 4.5 \ \mu\text{mhos}$$

The parameters which determine frequency response are

$$f_\alpha = \frac{\omega_\alpha}{2\pi} = \frac{D_n}{\pi w^2} = \frac{90}{\pi(3.6 \times 10^{-3})^2} = 2.45 \text{ mcps}$$

$$C_T \simeq 2400 \ \mu\mu fd/\text{cm}^2 \text{ for collector}$$

$$C_T \simeq 6500 \ \mu\mu fd/\text{cm}^2 \text{ for emitter}$$

so that

$$C_{T_c} = 6 \ \mu\mu fd$$

$$C_{T_e} = 16 \ \mu\mu fd$$

The electron storage capacitances of the emitter and collector diffusion admittances of this n-p-n transistor are:

$$C_{se} = \frac{g_{ee}}{1.5\omega_\alpha} = \frac{0.039}{1.5 \times 2.45 \times 2\pi \times 10^6} = 1690 \ \mu\mu fd$$

$$C_{sc} = \frac{g_{cc}}{1.5\omega_\alpha} = 0.2 \ \mu\mu fd$$

It is obvious that diffusion effects are most important at the emitter, while depletion layer capacitance is important at the collector.

The ohmic base spreading resistance of this unit is given (probably within a factor of two) by

$$r_b' \simeq \frac{\rho b}{w} = \frac{1.0}{3.6 \times 10^{-3}} = 278 \text{ ohms}$$

The low-frequencies values of the hybrid ("h") parameters may be calculated from the conductances, etc.. given above, but more direct methods are available and are used. The expressions required for both types of computation are given, however.

$$h_{21} \simeq -a_{ce} \simeq -0.985$$

$$h_{12} \simeq (g_{cc}/g_{ee}) + h_{22}r_b' + \mu_{bc}$$

$$\simeq \frac{kT}{qw}\frac{\partial w}{\partial V_c} \simeq 1.15 \times 10^{-4}$$

$$h_{22} \simeq \frac{g_{cc}g_{ee} - g_{ec}g_{ce}}{g_{ee}} = g_{cc}[1 - a_{ec}a_{ce}]$$

$$g_c' = h_{22} = \frac{I_e}{w}[2(1 - \beta_0) + (1 - \gamma_0)]\frac{\partial w}{\partial V_c} = 0.11\ \mu mho$$

$$h_{11} \simeq \frac{1}{g_{ee}} + (1 - a_{ce})r_b'$$

$$\simeq 26 + (0.015)278 \simeq 30.5 \text{ ohms}$$

These parameter values are in the range commonly encountered in drawn crystal units of good quality. It is obvious that this unit may be unstable at high temperatures, since the current transmission factor (alpha) is greater than unity at 70°C. The rise in alpha is the result of the very large increase in the hole density in the collector which accompanies the temperature rise.

In terms of the conventional tee network of r_e, r_b, r_c, and α, two points are interesting. First, the collector resistance $r_c \simeq 1/h_{22} \simeq 9$ megohms may seem very high. Actually, the lower values so commonly encountered are primarily the result of high leakage conductance, rather than a large electronic conductance. Second, the tee network base resistance $r_b \simeq h_{12}/h_{22} \simeq 1000$ ohms is nearly four times the high frequency feedback base resistance of 278 ohms.

5.3 p-n-p Fused Junction Transistor

The physical structure assumed for this unit is that of Fig. 2(b). The ring base contact is of particular importance. The material and structure are chosen to facilitate comparison with a specific development model for which a large amount of data is available. The emitter diameter is 15 mils and the collector diameter 30 mils. The base contact ring diameter is 40 mils.

The physical parameters of interest are:

V_c Collector voltage, -4.5 volts

I_e Emitter current, 1 ma

w_3 Wafer thickness, 3.5 mils

w_2 Collector to surface thickness, 2.1 mils

w_1 Collector to emitter thickness (not including depletion layers), 3.6×10^{-3} cm or 1.41 mil

ρ_b Base resistivity, 1.5 ohm cm

τ_b Hole lifetime in base region, 20 μ sec

ρ_e, ρ_c Emitter and Collector resistivities unknown but very low, probably 0.001 ohm cm

τ_e, τ_c Electron lifetime in emitter and collector unknown but very low, probably 0.1 μ sec

A_e Emitter area, 0.00114 cm^2

A_c Collector area, 0.00456 cm^2

The transmission parameters are then

$$g_{ee} = 0.039 \; mho$$

$$\gamma_0 \text{ probably} \geq 0.995$$

$$\beta_0 = 0.993$$

$$\alpha_0^* \simeq 1.0000$$

$$a_{ce} \geq 0.988$$

$$\frac{\partial w}{\partial V_c} = \frac{x_m}{2V_c} = \frac{2.6 \times 10^{-4}}{2 \times 4.5} = 2.89 \times 10^{-5} \text{ cm/volt}$$

$$g_{cc} \simeq 8.15 \; \mu mhos$$

The parameters which determine frequency response are:

$$f_\alpha = \frac{\omega_\alpha}{2\pi} = \frac{D_p}{\pi w_1^2} = \frac{44}{\pi(3.6 \times 10^{-3})^2} = 1.08 \; mcps$$

$$C_T = 5{,}000 \; \mu\mu fd/cm^2 \text{ for collector}$$

$$C_T \simeq 20{,}000 \; \mu\mu fd/cm^2 \text{ for emitter}$$

so that

$$C_{Tc} = 22.8 \; \mu\mu fd$$

$$C_{Te} = 22.8 \; \mu\mu fd$$

The effective hole storage capacitances of the emitter and collector are:

$$C_{se} = 3840\ \mu\mu fd$$

$$C_{sc} = 0.8\ \mu\mu fd$$

By the equations given in Section 3.0

$$r_b'_1 = 55 \text{ ohms}$$

$$r_b'_2 = 35 \text{ ohms}$$

The low frequency values of the "h" parameters are

$$h_{21} \simeq -0.988$$

$$h_{12} \simeq 2.09 \times 10^{-4}$$

$$h_{22} \simeq 0.17\ \mu mho$$

$$h_{11} \simeq 26 + (0.012)55 \simeq 26.7$$

These theoretical values are compared with observed values in Table I. The major discrepancy in h_{21} is charged to surface recombination of injected holes, which was ignored in the calculation. It should be noted that h_{22} becomes 0.45×10^{-6} *mho* if the calculated value is corrected by the ratio 0.032/.012, which is the ratio of the measured and calculated $(1 - \alpha)$'s or $(1 + h_{21})$'s. The difference between computed and measured h_{11} is the sum of a number of effects. First, the actual $(1 - \alpha)$ is greater than the computed value. Next, the junction temperature was probably greater than the assumed 25°C. Finally, the carrier injection level is high enough to modify the emitter diode properties in this direction.

The difference between calculated and observed current transmission cutoff is greater than appears from the data, since the theoretical three db response frequency is about 22 per cent higher than the "alpha

TABLE I

	Calculated	Measured
h_{21}	− 0.988	−0.968
h_{22}	0.17×10^{-6} mho	0.48×10^{-6} mho
h_{12}	2.09×10^{-4}	1.85×10^{-4}
h_{11}	26.7 ohm	33 ohm
C_c	22.8 $\mu\mu fd$	24.7 $\mu\mu fd$
f_α	1.08 mcps	0.95 mcps
$r_b'_1$	55 ohm	55 ohm
$r_b'_2$	35 ohm	63 ohm

cutoff" frequency calculated here. The difference is believed to be the result of the fact that holes emitted around the emitter periphery have much longer transit paths than do those emitted into the region directly between the electrodes and consequently reduce significantly the current cutoff frequency.

The serious discrepancy in $r_b'_2$ is probably the result of the $r_b'_1$ calculation being very pessimistic because of neglect of peripheral emission effects and of w_2 and w_3 being somewhat smaller than the assumed values. Again it can be seen that the equivalent tee base resistance $r_b = h_{12}/h_{2_2}$ is 375 ohms, nearly seven times the high frequency resistance of 55 ohms.

5.4 Qualitative Comparison

As might be expected, the qualitative agreement between theory and observation is better than the quantitative. For example, C_c, h_{22}, and h_{12} vary approximately as $(V_c')^{-1/2}$ in fused junction units. Alpha cutoff frequency increases as collector reverse bias is increased—a natural result of the narrowing of the base region. The qualitative discrepancies that are found are usually associated with large experimental deviations from the assumptions of the analysis.

5.5 Review of Design Calculations

Numerical analysis of both drawn junction and fused junction transistors has shown rather good agreement of theory and experiment. It is necessary, however, to modify some of the results empirically because of lack of full understanding of some effects, such as leakage and surface recombination.

Qualitative agreement of measurements and theory is, of course, much better than the quantitative correlation. For example, the dependence of all of the parameters on emitter current and collector voltage is almost exactly that expected from theory. Some of the dependence on emitter currents involves high carrier injection level theory which has been omitted from this study.

6.0 SUMMARY

6.1 Transmission Theory

Design theory of the small signal transmission parameters of junction transistors is relatively complete.

A one-dimensional analysis of minority carrier diffusion currents in terms of short circuit admittances has been combined with a similar analysis of depletion layer capacitances and an approximate three-

dimensional analysis of ohmic base region spreading resistance. The resulting equivalent circuit has characteristics in good agreement with experimental observations. In particular, collector capacitance, ohmic base region spreading resistance and the current transmission three db cutoff frequency may be computed with fair accuracy. The low frequency values of the common base hybrid parameters may also be calculated, but neglect of surface recombination and surface leakage results in serious errors in the short circuit current transmission factor h_{21} and the open circuit collector conductance h_{22}. The deviations of these two parameters from calculated values are, however, both reasonable and mutually consistent. The ohmic base layer spreading resistance, *which is the only base resistance of importance at high frequencies*, is very often much smaller than the low frequency base resistance appearing in an equivalent tee network.

Qualitative agreement of theory and measurement is excellent. The variation of all parameters with emitter current and collector voltage is within a few per cent that predicted from theory.

6.2 State of the Art

Since this paper was deliberately limited in scope, it is pertinent both to review its objectives and to point out significant omissions. The principal objective sought was presentation of small signal transmission design theory. No attempt was made to give a simple explanation of the junction transistor, relating both its large signal and its transmission characteristics to simple physical assumptions. While the design theory presented consolidates in one place some already published information, much remains to be done in assembling and integrating such knowledge from its present widely scattered locations.

In addition there exists a more detailed understanding of junction transistor characteristics than can be found in the literature. For example, units are found occasionally with negative h_{12}. This is a result of an easily modulated high resistance between base region and base contact (high μ_{bc}). Publication of such information can reduce by a few db the amount of head-scratching done by production engineers.

Other phenomena for which explanations have been developed are surface recombination and high carrier injection level effects. Despite this, much work remains to be done.

ACKNOWLEDGEMENTS

The general point of view taken here has been much influenced by discussions with J. A. Morton. Comments and criticisms by R. M. Ryder have been particularly helpful in the preparation of this material.

APPENDIX A

1.0 GENERAL

This study is an extension of Shockley's analysis of the junction transistor to include high-frequency effects and the voltage dependence of base-layer thickness. Shockley's paper* and the later paper by Shockley, Sparks, and Teal† contain the following of interest here:

(a) analysis of the dc steady state of a junction transistor*;
(b) analysis of the low-frequency small-signal parameters r_e, α, C_c*†; and
(c) analysis of frequency dependence of the transport factor β†.

In addition to repeating the above, this study gives these new results:

(a) analysis of steady-state small-signal ac operation [dc biases present]; and
(b) the small-signal ac short-circuit admittances y_{ee}, y_{ce}, y_{ec}, and y_{cc}.

1.1 ASSUMPTIONS

Semiconductor.

The p-n-p type structure assumed is shown in Fig. 6. The emitter, base, and collector regions may each be characterized by a resistivity and a minority carrier diffusion length. The emitter, collector, and base contacts have no effect on the currents which flow at the junctions. Injected carriers pass through the base layer by diffusion and through

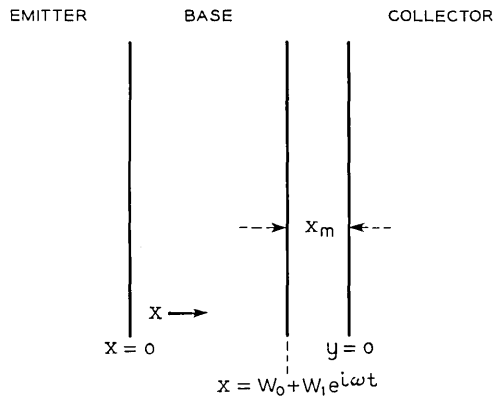


Fig. 6 — p-n-p junction transistor.

* W. Shockley, The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors, B.S.T.J., **28**, p. 435.

† W. Shockley, M. Sparks, and G. K. Teal, The p-n Junction Transistors, Phys. Rev., **83**, p. 151, July, 1951.

the barrier layers* by drift. Base-layer thickness w depends on collector potential V_c through variation of the collector barrier thickness x_m. Variations of emitter barrier thickness are unimportant. The junctions are parallel.

Currents and Potentials

The currents and potentials studied are those at the collector and emitter barriers. Unless otherwise specified, "potential" implies difference in majority carrier Fermi levels, i.e., externally applied potential, rather than difference in electrostatic potential. The collector reverse potential is assumed to be many multiples of kT/q(e.g. $>$ 0.5 volts), so that the classical p-n diode reverse conductance may be neglected. At 0.5 volt, this is of the order of 10^{-11} mhos. It decreases in magnitude one decade per sixty millivolts of bias potential.

Base Resistance

Majority carrier resistance in the base layer is not considered here.

Other Assumptions

Surface effects are excluded from consideration. In addition, several mathematical approximations of little physical consequence appear in the text as needed.

1.2 METHOD

The procedure employed is substantially that used by Shockley with some additions. First, minority carrier concentrations on both sides of each barrier are related to the barrier potentials.

The dc minority carrier distribution in each of the three transistor regions is then computed from these boundary conditions with the aid of the continuity equation.

Next, small-signal ac perturbations of the barrier potentials and of the minority carrier densities at the barriers are used in the same way to find ac distributions of the minority carriers. The effects of voltage dependence of base-layer thickness are found by means of a small-signal ac perturbation of the position of the collector side of the base layer. The resulting ac distribution of minority carriers is computed as before.

Finally, dc and ac currents at emitter and collector barriers are com-

* I.e., charge depletion layers.

puted from the gradients of the minority carrier densities at the barriers.

In the ac analysis, the forward rotating time function $e^{i\omega t}$ is used. In general, the first-order solution for small signals is obtained by assuming that the disturbance associated with $e^{i\omega t}$ is small, by neglecting its powers and harmonics, and by using first-order expansions, e.g.,

$$e^{\frac{w_0}{L}+\frac{w_1}{L}e^{i\omega t}} \approx \left(1+\frac{w_1}{L}e^{i\omega t}\right)e^{\frac{w_0}{L}}$$

The ac magnitudes such as w_1, p_{e1}, n_{c1}, represent complex phasors of the form $ae^{j\phi}$. A list of symbols is given in Section 1.5.

1.3 ANALYSIS

In the base layer at $x = 0$,

$$p = p_{e0} + p_{e1}e^{i\omega t} = p_n e^{qV_e/kT} \tag{1a}$$

at $x = w_0$,

$$p = p_{c0}+p_{c1}e^{i\omega t} = p_n e^{qV_c/kT} \tag{1b}$$

in which

$$V_e = V_{e0} + V_{e1}e^{i\omega t}$$

$$V_c = V_{c0} + V_{c1}e^{i\omega t}$$

and $V_{e1} \ll V_{e0}$ and $V_{c1} \ll V_{c0}$ so that

$$p_{e0} \approx p_n e^{qV_{e0}/kT} p_{c1} \approx p_{e0}\frac{q}{kT}V_{c1}$$

$$p_{co} \approx p_n e^{qV_{c0}/L} p_{c1} \approx p_{co}\frac{q}{kT}V_{c1}$$

P_{e0}, P_{e_0}, V_{e_0}, and V_{c0} are average values while p_{e1}, p_{c1}, V_{e1}, and V_{c1} are ac phasors.

The continuity equation for holes for one-dimensional flow is:

$$D_p\frac{\partial^2 p}{\partial x^2} - \left(\frac{p_n - p}{\tau}\right) = \frac{\partial p}{\partial t} \tag{2}$$

in which τ is hole lifetime. A solution to equation (8–18) is:

$$p = p_n + Ae^{x/L} + Be^{-x/L} + Ce^{sx/L+i\omega t} + De^{-sx/L+i\omega t} \tag{3}$$

where $L = \sqrt{D\tau}$ and $s = (1 + i\omega\tau)^{1/2}$.

Application of the boundary values of equation (1a) and (1b) to

equation (3) gives hole density in the base layer:

$$\begin{aligned} p_0(t,x) = p_n &+ \left[\frac{(p_{c0} - p_n) - (p_{e0} - p_n)e^{-w_0/L}}{2 \sinh (w_0/L)}\right] e^{x/L} \\ &- \left[\frac{(p_{c0} - p_n) - (p_{e0} - p_n)e^{w_0/L}}{2 \sinh (w_0/L)}\right] e^{-x/L} \\ &+ \left[\frac{p_{c1} - p_{e1}e^{-sw_0/L}}{2 \sinh (sw_0/L)}\right] e^{sx/L + i\omega t} \\ &- \left[\frac{p_{c1} - p_{e1}e^{sw_0/L}}{2 \sinh (sw_0/L)}\right] e^{-sx/L + i\omega t} \end{aligned} \tag{4}$$

Since up to this point w has been assumed constant [$w = w_0$], equation (4) does not include effects of voltage dependence of base layer thickness. To introduce these, a new set of boundary conditions is used:
at $x = 0$,

$$p = p_{e0} + p_{e1}e^{i\omega t} \tag{5a}$$

and at $x = w_0 + w_1 e^{i\omega t}$,

$$p = p_{c0} + p_{c1}e^{i\omega t} \tag{5b}$$

in which $w_1 \leqq w_0$ and is a phasor.

$$w_1 = \frac{\partial w}{\partial V_c} V_{c1}$$

It can be seen that conditions are as before except that the collector side of the base layer swings about position w_o at angular frequency ω.

A solution of equation (31) with conditions (5a) and (5b) is given by

$$p_1(t,x) = p_0(t,x) + \tilde{p}(t,x) \tag{6}$$

in which $p_0(t,x)$ is given by equation (4) and $\tilde{p}(t,x)$ is the perturbation associated with $w_1 e^{i\omega t}$.

If equations (5a) and (5b) are rewritten in terms of $\tilde{p}(t,x)$, they become, using first order expansions:

$$\tilde{p}(t,o) = 0 \tag{7a}$$

$$\begin{aligned} \tilde{p}(t,w_0 + w_1 e^{i\omega t}) = [&(p_{e0} - p_n) \operatorname{csch} (w_0/L) \\ &- (p_{c0} - p_n) \coth (w_0/L)] \frac{w_1}{L} e^{i\omega t} \end{aligned} \tag{7b}$$

Since $\tilde{p}(t,x)$ is an ac solution of the continuity equation (2), it has the form:

$$\tilde{p}(t,x) = Ee^{sx/L + i\omega t} + Fe^{-sx/L + i\omega t} \tag{8}$$

Use of equations (7a) and (7b) leads to:

$$p(t,x) = \frac{\sinh (sx/L)}{\sinh (sw_0/L)} [(p_{e0} - p_n) \operatorname{csch} (w_0/L) - (p_{c0} - p_n) \coth (w_0/L)] \frac{w_1}{L} e^{i\omega t} \tag{9}$$

The complete solution for hole density in the base layer is:

$$\begin{aligned} p_1(t,x) = p_n &+ \left[\frac{(p_{c0} - p_n) - (p_{e0} - p_n)e^{-w_0/L}}{2 \sinh (w_0/L)}\right] e^{x/L} \\ &- \left[\frac{(p_{c0} - p_n) - (p_{e0} - p_n)e^{w_0/L}}{2 \sinh (w_0/L)}\right] e^{-x/L} \\ &+ \left[\frac{p_{c1} - p_{e1}e^{-sw_0/L}}{2 \sinh (sw_0/L)}\right] e^{sx/L + i\omega t} \\ &- \left[\frac{p_{c1} - p_{e1}e^{sw_0/L}}{2 \sinh (sw_0/L)}\right] e^{-sx/L + i\omega t} \\ &+ \frac{w_1}{L} e^{i\omega t} \frac{\sinh (sx/L)}{\sinh (sw_0/L)} [(p_{e0} - p_n) \operatorname{csch} (w_0/L) \\ &\qquad - (p_{c0} - p_n) \coth (w_0/L)] \end{aligned} \tag{10}$$

The hole-current density in the base layer is found from equation (10) by the use of the equation for diffusion current

$$I_p = -qD_p \frac{dp}{dx} \tag{11}$$

which yields

$$\begin{aligned} I_p = -q \frac{Dp}{L} \Bigg(&(p_{c0} - p_n) \frac{\cosh (x/L)}{\sinh (w_0/L)} - (p_{e0} - p_n) \cdot \\ &\frac{\cosh [(x - w_0)/L]}{\sinh (w_0/L)} + sp_{c1} \frac{\cosh (sx/L)}{\sinh (sw_0/L)} e^{i\omega t} \\ &- sp_{e1} \frac{\cosh [(sx - sw_0)/L]}{\sinh (sw_0/L)} e^{i\omega t} + s \frac{w_1}{L} e^{i\omega t} \frac{\cosh (sx/L)}{\sinh (sw_0/L)} \\ &[(p_{e0} - p_n) \operatorname{csch} (w_0/L) - (p_{c0} - p_n) \coth (w_0/L)] \end{aligned}$$

Hole-current densities at emitter and collector may be found by substitution of $x = 0$ and $x = w_0$ respectively. The first two terms in equation (12) give dc current components* which may be attributed to

* With some labor, these terms may be converted to Shockley's equation (5.6).

collector and emitter potentials in that order, while the last three are ac terms resulting from collector potential, emitter potential, and base layer thickness variation respectively.

The hole-current densities may be related to the ac potentials by use by use of the approximations given at the beginning of paragraph 1.3.

$$p_{e1} \approx p_{e0} \frac{q}{kT} V_{e1} \qquad p_{c1} \approx p_{c0} \frac{q}{kT} V_{c1}$$

and the relation

$$w_1 = \frac{\partial w}{\partial V_c} V_{c1}$$

However, the average collector potential V_{c0} is negative and many times kT/q so that $p_{c0} \simeq 0$ and $pc_1 = 0$. The ac emitter hole-current density is therefore

$$\begin{aligned} I_{pe1}e^{i\omega t} = \frac{qDp}{L}\Bigg(& V_{e1} \frac{q}{kT} sp_{e0} \coth (sw_0/L) \\ & - \frac{V_{c1}}{L} \frac{\partial w}{\partial V_c} s \operatorname{csch} (sw_0/L)[(p_{e1} - p_n) \operatorname{csch} (w_0/L) \\ & - (q_{c0} - p_n) \coth (w_0/L)]\Bigg)e^{i\omega t} \end{aligned} \tag{13a}$$

If $p_{e0} > p_n$, equations (13a) and (13b) becomes very closely

$$I_{pe1} = \left[I_{pe0} \frac{q}{kT} \frac{s \tanh (w_0/L)}{\tanh (sw_0/L)} V_{e1} + I_{pc0} \frac{\partial w}{\partial V_c} \cdot \frac{sV_{c1}}{L \sinh (sw_0/L)} \right] \tag{13b}$$

In equation (13b), I_{pe0} and I_{pc0} are average emitter and collector hole-current densities. The entire coefficients of V_{e1} and V_{c1} in equation (13b) are the input and the feedback short-circuit admittances associated with hole flow in the transistor.

Similarly, forward transfer and output short-circuit admittances associated with hole flow may be found from equation (12) by stubstitution of $x = w_0$ and use of the approximation $p_1 \simeq p_1 - p_n$, i.e., $p_1 \gg p_n$. In calculating collector current, the sign of equation (12) must be reversed, since equation (12) gives current flow in the x-direction while collector current is assumed to flow in the negative x-direction. The admittances are given in the summary at the end.

Next, there are admittances associated with electron flow in the p-n-p transistor. Flow of electrons from base to emitter gives rise to an input admittance term, while electrons flowing from collector to base give rise to output and forward transfer admittance terms. An outline

of the derivation of these terms follows. The terms are given in the summary.

For electrons in the emitter:

at $x = 0$,

$$n_e = n_{e0} + n_{e1}e^{i\omega t} = n_p e^{qV_e/kt}$$

at $x = -\infty$,

$$n_e = n_p$$

in which, again, $V_e = V_{e0} + V_{e1}e^{i\omega t}$ and first-order expansions are used. $x = -\infty$, $n_e = n_p$, is chosen as a boundary condition in order to eliminate effects of the emitter contact. Solution of the continuity equation results in:

$$n_{0e}(t,x) = n_p + (n_{0e} - n_p)\, e^{x/L} + n_{01}e^{-sx/L+i\omega t} \tag{14}$$

in which the τ is $s = (1 + i\omega\tau)^{1/2}$ now implies electron lifetime in the emitter body. Electron diffusion current at the emitter is computed by means of the equation corresponding to equation (11), giving

$$I_{ne} = \frac{qD_n}{L_n}((n_{e0} - n_p) + n_{e1}e^{i\omega t} \tag{15}$$

The ac admittance associated with the last term appears in the summary, Section 1.4.

The ac electron current from the collector is not a diffusion current, but rather a drift current resulting from the hole current flowing in the collector body. Since the ac electron current is directly proportional to the ac hole current in the collector, the result is an effective multiplication of the output and forward transfer admittances associated with hole current in the collector body.*

For electrons in the collector the boundary conditions are:

at $y = 0$,

$$n = n_p e^{qV_c/kT} = 0 \tag{16a}$$

at $y = \infty$,

$$n = n_p \tag{16b}$$

The condition $n = 0$ at the edge of the barrier region results from V_c

* Space-charge layer widening effects are neglected since they are usually very small and are difficult to analyze.

being negative and many times kT/q. The distribution of electrons in the collector is governed by conditions (16a) and (16b) and the modified continuity equation* developed by W. van Roosbroeck.

$$\frac{\partial n}{\partial t} = -\frac{n - n_p}{\tau} + \frac{\mu_n E_a}{M_0}\frac{\partial n}{\partial y} + D_0\frac{\partial^2 n}{\partial y^2} \tag{17}$$

in which E_a is the electric field which would be associated with the total current if $n \equiv n_p$. $M_0 = (p_p + bn_0)/(p_p + n_p)$ and $D_0 = (p_p + n_p)/(p_p/D_n + n_p/D_p)$.

If n is assumed to be given by $n(t,x) = n_0(y) + n_1(y)e^{i\omega t}$ and E is assumed to be

$$E_a(t,x) = E_0 + E_1 e^{i\omega t} = \rho_c I_{c0} + \rho_{pc} I_{pc1} e^{i\omega t}$$

in which I_{c0} is average collector current and I_{pc1} is the ac hole current of the collector, equation (17) may be reduced to two equations

$$0 = -\left(\frac{n_0 - n_p}{D_0\tau}\right) + \frac{\mu_n E_0}{D_0 M_0}\frac{\partial n_0}{\partial y} + \frac{\partial^2 n_0}{\partial y^2} \tag{18a}$$

$$-\frac{\mu_n E_1}{M_0 D_0}\frac{\partial n_0}{\partial y} = -\frac{1 + i\omega\tau}{D_0\tau} n_1 + \frac{\mu_n E_0}{D_0 M_0}\frac{\partial n_1}{\partial y} + \frac{\partial^2 n_1}{\partial y^2} \tag{18b}$$

The solution of equation (18a) under condtiions (16a) and (16b) is

$$n_0 = n_p(1 - e^{m_2 y}) \tag{19}$$

where

$$m_2 = -\frac{\mu_n E_0}{2D_0 M_0} - \sqrt{\left(\frac{\mu_n E_0}{2D_0 M_0}\right)^2 + \frac{1}{D_0\tau}}$$

Inspection of equation (19) shows that

$$\frac{\partial n_0}{\partial y} = -m_2 n_p e^{m_2 y} \tag{20}$$

It is then apparent that n_1 must be of the form

$$n_1 = n_p e^{m_2 y} f(y) \tag{21}$$

Substitution of equation (2) and equation (21) into (18b) leads to

$$\frac{\partial^2 f}{\partial y^2} + \left(m_2 + \frac{\mu_n E_0}{M_0 D_0}\right)\frac{\partial f}{\partial y} - \frac{i\omega}{D_0} f = \frac{n_p m_2 \mu_n e_2}{M_0 D_0} \tag{22}$$

A general solution of equation (22) is

$$f(y) = \frac{n_p m_2 \mu_n E_1}{-i\omega M_0} + Be^{r_1 y} + Ce^{r_2 y} \tag{23}$$

* W. van Roosbroeck; private communication.

where

$$r_1 \cdot r_2 = -\frac{1}{2}\left(m_2 + \frac{\mu_n E_0}{D_0 M_0}\right) \pm \sqrt{\left[\frac{1}{2}\left(m_2 + \frac{\mu_n E_0}{D_0 M_0}\right)\right]^2 + \frac{i\omega}{D_0}}$$

It may be shown that the boundary conditions on $f(y)$ are:

$$\text{at } y = 0, \quad f = 0; \qquad \text{at } y = \infty, \quad f = 0.$$

Application of these values to equation (23) results in

$$f(y) = \frac{n_p m_2 \mu_n E_1}{i\omega M_0}\left[1 - e^{r_2 y}\right] \tag{24}$$

in which r_2 is obtained using the negative square root.

The electron density in the collector is now given by

$$n = n_p(1 - e^{m_2 y}) + \frac{n_p m_2 \mu_n E_1}{i\omega M_0}\left[1 - e^{r_2 y}\right]e^{m_2 y} e^{i\omega t} \tag{25}$$

The electron current is given by

$$I_n = n_p q D_n \left[-m_2 e^{m_2 y} + e^{i\omega t}\frac{\mu_n E_1 m_2}{i\omega M_0}(m_2 - m_2 e^{r_2 y} - r_2 e^{r_2 y})e^{m_2 y}\right]. \tag{26}$$

At the collector junction, the ac component reduces to

$$I_{nc1} = \frac{n_p q D_n \mu_n m_2 r_2 E_1}{i\omega M_0} \tag{27}$$

Now, since $E_1 = \rho_{pc} I_{pc1}$, the collector multiplication factor $(I_{pc1} + I_{nc1})/I_{pc1}$ is

$$\alpha^* = 1 + \frac{n_p q D_n \mu_n m_2 r_2}{i\omega M_0}\rho_{pc} \tag{28a}$$

$$\alpha^* = 1 + \frac{\sigma_{nc}}{\sigma_{pc}}\frac{D_n m_2 r_2}{i\omega M_0} \tag{28b}$$

The effect of collector multiplication as given in equations (28a) and (28b) is included in the general admittance expressions given later.

Finally, no mention has been made of the admittances associated with barrier capacitances. Since the currents which charge these are majority carrier currents, there are no input-output interactions except those associated with majority carrier resistance of the base layer [an effect not analyzed in this study]. These capacitances add directly to input and output admittances. Shockley* gives methods for calculating these capacitances.

* Loc. cit., vol. 28, page 435.

1.4 SUMMARY

In paragraph 1.3, derivations were given or outlined for each of the four small-signal short-circuit admittances associated with the hole flow, electron flow, and barrier capacitances in junction transistors. The terms appear in that order in the expressions which follow.

$$y_{ee} = \left[\frac{qp_nD_ps_p}{L_p}\coth\ (s_pw_0/L_p) + qn_pD_ns_{ne}/L_{ne}\right]\frac{q}{kT}\,e^{qV_{e0}/kT} + i\omega C_{Te} \tag{29}$$

$$y_{ce} = -\left[q\,\frac{p_nD_ps_p}{L_p}\,\mathrm{csch}\ (s_pw_0/L_0)\right]\frac{q}{kT}\,e^{qV_{e0}/kT}\left(1 + \frac{\sigma_{nc}}{\sigma_{pc}}\,\frac{D_nm_2r_2}{i\omega M_0}\right) \tag{30}$$

$$y_{ec} = -\frac{\partial w}{\partial V_c}\,\frac{s_p}{L_p\sinh\ (s_pw_0/L_p)}\,\frac{qp_nD_p}{L_p}\,[(e^{qV_{e0}/kT} - 1)\cdot \mathrm{csch}\ (w_0/L_p) + \coth\ (w_0/L_p)] \tag{31}$$

$$y_{cc} = \left(\frac{\partial w}{\partial V_c}\,\frac{s_p}{L_p\tanh\ (s_pw_0/L_p)}\,\frac{qp_nD_p}{L_p}\,[(e^{qV_{e0}/kT} - 1)\cdot \mathrm{csch}\ (w_0/L_p) + \coth\ (w_0/L_p)] + i\omega C_{Te}\right)\left(1 + \frac{\sigma_{nc}D_nm_2r_2}{\sigma_{pc}i\omega M_0}\right) \tag{32}$$

in which all symbols are defined in Section 1.5. It should be noted that collector multiplication operates on the current to the barrier capacitance since the latter current is a hole current in the collector body.

The term $\partial w/\partial V_c$ is the same for both p-n-p and n-p-n structures. It is: for step junctions

$$\frac{\partial w}{\partial V_c} = -\frac{x_m}{2V_c} \tag{33}$$

and for graded junctions

$$\frac{\partial w}{\partial V_c} = -\frac{x_m}{6V_c} \tag{34}$$

V_c in equations (33) and (34) means dc electrostatic potential difference across the collector barrier.

Equations (29) through (32) may be manipulated into many forms. One of these sets which may be employed as a starting point for the approximate forms given in the body of this chapter is:

$$y_{ee} = \frac{q}{kT}\left[I_{pe0}\,\frac{s_p\tanh\ (w_0/L_p)}{\tanh\ (s_pw_0/L_p} + I_{ne}s_{se}\right] + i\omega C_{Te} \tag{35}$$

$$y_{ce} = -\frac{q}{kT} I_{pe0} \frac{s_p \tanh (w_0/L_p)}{\sinh (s_p w_0/L_p)} \left[1 + \frac{\sigma_{nc}}{\sigma_{pc}}\right]^* \tag{36}$$

$$x_{cc} = \frac{\partial w}{\partial V_c} \frac{s_p}{L_p \sinh (s_p w_0/L_p)} I_{pc0} \tag{37}$$

$$y_{cc} = \left[\frac{-\partial w}{\partial V_c} \frac{s_p}{L_p \tanh (s_p w_0/L_p)} I_{pc0} + i\omega C_{Te}\right] \left[1 + \frac{\sigma_{nc}}{\sigma_{pc}}\right]^* \tag{38}$$

The change in signs which occurs in going from equations (31) and (32) to equations (37) and (38) takes place because the current replaced by I_{pc} had the opposite assigned positive sense.

1.5 SYMBOLS USED IN THE APPENDIX

C_c , C_{Tc} = collector barrier capacitance.
C_{Te} = emitter barrier capacitance.
D_n , D_p = diffusion constants for electrons and holes
$D_0 = (p_p + n_p)/(p_p/D_n + n_p/D_p)$
$E_a = i\rho$ = electric field associated with current at thermal equilibrium carrier densities.
E_1 , E_0 = ac and dc components of E_a
I_{ne} , I_{ne0} , I_{ne1} , I_{nc} , I_{nc0} , I_{nc1} = total, average, and ac emitter and collector electron currents
I_{pe} , I_{pe0} , I_{pe1} , I_{pc} , I_{pc0} , I_{pc1} = total, average, and ac emitter and collector hole currents
kT/q = thermal energy of carriers = 0.026 electron-volt
L_p , L_n = diffusion lengths for holes and electrons
$m_2 = -\frac{\mu_n E_0}{2D_0 M_0} - \sqrt{\left(\frac{\mu_n E_0}{2D_0 M_0}\right)^2 + \frac{1}{D_0 \tau}}$ decay constant for average electron density in the collector
$M_0 = (p_p - bn_p)/(p_p = n_p)$
n_p , n_n = thermal equilibrium electron densities in p and n regions
n_{e0} , n_{e1} = dc and ac components of electron density at emitter junction
p_n, p_p = thermal equilibrium hole densities in n and p regions
p_{e0} , p_{e1} , p_{c0} , p_{c1} = dc and ac components of hole density at emitter and collector junctions
q = electronic charge, 1.6×10^{-19} coulombs
r_e = ac emitter resistance
$r_1, r_2 = -\frac{1}{2}\left(m_2 + \frac{\mu_n E_0}{D_0 M_0}\right) \pm \sqrt{\left[\frac{1}{2}\left(m_2 + \frac{\mu_n E_0}{D_0 M_0}\right)\right]^2 + \frac{i\omega}{D_0}}$

* The factor $a^* = (1 + \sigma_c/\sigma_{pc})$ is current dependent. At small average collector currents, it is $(1 + \sigma_{nc}/2\sigma_{pc})$ and rises to $(1 + \sigma_{nc}/\sigma_{pc})$ at high current densities.

$s = (1 + i\omega\tau)$
T = temperature in degrees Kelvin (absolute temperature)
V_e, V_{e0}, V_{e1}, V_c, V_{c0}, V_{c1} = total, average, and ac emitter and collector potentials.
V_c = average collector potential
w, w_0, w_1 = total, average, and ac variation of base-layer thickness
x = distance from emitter barrier
x_m = thickness of collector barrier layer
y = distance from collector side of collector barrier layer
y_{ee}, y_{ce}, y_{ec}, y_{cc} = short-circuit ac admittances
α = current amplication factor
β = base-layer transport factor
ρ_c = resistivity of collector region
ρ_{pc} = resistivity for holes in collector region
σ_{pc}, σ_{nc} = conductivities for holes and electrons in collector region
τ = minority carrier lifetime
μ_n = electron mobility
μ_p = hole mobility
ω = angular frequency in radians

Appendix B

Base Layer Spreading Resistance

Significance

The bulk resistance of the base layer or base layer spreading resistance (r_b') is important because base current passing through r_b' produces a base contact to emitter junction negative feedback voltage $i_b r_b'$. Reduction of r_b' is an important objective in improvement of junction units.

Types

Since feedback voltage to the emitter junction is produced by two separately measurable base current components having very different flow paths, two separately measurable base layer spreading resistances ($r_b'_1$ and $r_b'_2$) may be defined.

The current $(1\text{-}\alpha)i_e$ originates in the base layer between the emitter and collector junctions and flows radially through the base layer to the base contact producing a feedback voltage at the emitter; $r_b'_1$ is defined as the ratio of this feedback voltage to the current.

The collector capacitance current $j\omega C_c v_c$ enters the base layer uni-

formly over the entire area of the collector junction resulting in a feedback voltage at the emitter; $r_b'_2$ is defined as the ratio of this feedback voltage to this current.

Calculations

The resistances $r_b'_1$ and $r_b'_2$ for the transistor of Fig. 2(*b*) may be computed with the help of three formulas which give the feedback voltages for the three geometrical problems involved in this transistor. Each expression gives the voltage V developed at electrode C by a current I entering through electrode A and leaving through electrode B. The formulas are in terms of sheet resistance ρ_b/w (resistance per square in ohms) and the radii involved.

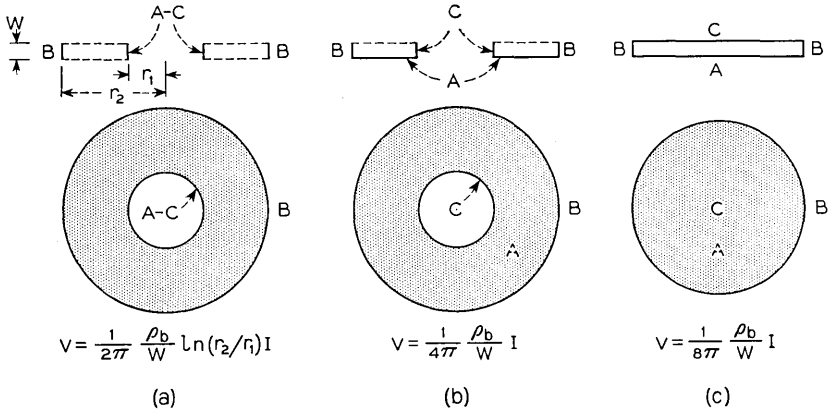


Fig. 7 — Feedback voltage for three geometries.

The simplest situation and its formula are shown in Fig. 7(a). Electrodes A and C are the same and expression gives the resistance of an annular ring to radial current flow.

Fig. 7(b) also shows an annular ring, but the current I is introduced uniformly over one of the flat surfaces, while the voltage V is measured from outside edge to inside edge.

Fig. 7(c) shows current introduced uniformly over the surface of a disc, while voltage V is the average voltage developed along the surface of the disc. It should be clear that electrodes A and C do not conduct parallel to the disc surface (i.e., they are not equipotentials).

Example

The $r_b's$ for the transistor of Fig. 2(b) will be calculated.

A. $r_b'_1$ — The current I = $(1\text{-}\alpha)i_e$ is assumed to originate uniformly

in the region between emitter and collector in Fig. 2(b). It flows radially outward to the base contact.

$$V = I\left[\frac{1}{8\pi}(\rho_b/w_1) + \frac{1}{2\pi}(\rho_b/w_2)\ln(r_2/r_1) + \frac{1}{2\pi}(\rho_b/w_3)\ln(r_3/r_2)\right]$$

$$r_{b1}' = V/I = \rho_b\left[\frac{1}{8\pi w_1} + \frac{1}{2\pi w_2}\ln\frac{r_2}{r_1} + \ln\left(\frac{r_3}{r_2}\right)\right]$$

B. $r_b'_2$ — Inspection of Fig. 7 shows that the collector capacitance current originating opposite the emitter junction (I_1) has the same flow path as the $(1\text{-}\alpha)i_e$ current. The remainder of the collector current which enters the base from the collector outside the emitter (I_2) likewise flows radially out to the base contact. The total feedback voltage developed between emitter junction and base contact is

$$V = I_2 r_{b1}' + I_2\left[\frac{1}{4\pi}(\rho_b/w_2) + \frac{1}{2\pi}(\rho_b/w_3)\ln(r_3 r_2)\right]$$

$$r_{b2}' = \frac{V}{1_1 + 1_2} = \rho\left\{\frac{1}{2\pi w_3}\ln\left\{\frac{r_3}{r_2}\right\} + \left\{\frac{r_1}{r_2}\right\}^2\left[\frac{1}{2\pi w_2}\ln\left\{\frac{r_2}{r_1}\right\} + \frac{1}{8\pi w_1}\right] + \left[1 - \left\{\frac{r_1}{r_2}\right\}^2\right]\frac{1}{4\pi w_2}\right\}$$

C. Numerical Results

For $\rho_b = 1.5$ ohm-cm

$w_1 = 0.5$ mil

$w_2 = 1.0$ mil

$w_3 = 2.0$ mil

$r_2/r_1 = 2$

$r_3/r_2 = 2$

$r_b'_1 = 147$ ohms

$r_b'_2 = 96$ ohms

Transistor Oscillator for Use in Multifrequency Pulsing Current Supply

By F. E. BLOUNT

(Manuscript received June 30, 1953)

This paper covers the design and performance of an oscillator using a transistor, in the multifrequency pulsing of digital information over telephone transmission media. The frequencies used are in the range from 700 to 1700 cycles per second.

INTRODUCTION

A large proportion of the telephone calls made require connections to be set up in more than one central office. Where common control systems are used, this requires that information needed by the second central office be transmitted to it from the first central office which in turn has received its information from the calling subscriber. The "language" used in some cases for transmitting the information is in the form of short pulses of alternating current. Each pulse consists of a combination of two of six available frequencies. Twelve combinations of the six frequencies make up the total "vocabulary". Ten are required for digits and two for special signals sent at the beginning and end of pulsing. This is known as multifrequency, MF, pulsing. The device which controls the pulses is called an MF sender. The device which receives the MF pulses and translates the information received for use by other equipment is called an MF receiver.

An operator may also use multifrequency pulsing when transmitting information by means of a key set to a distant office. Control of the frequencies used is obtained by contacts on the key depressed. Twelve keys are used.

The previous source of ac was a circuit capable of supplying both operator positions and senders. This equipment is however quite expensive. If only a few senders in an office require the MF current the cost per

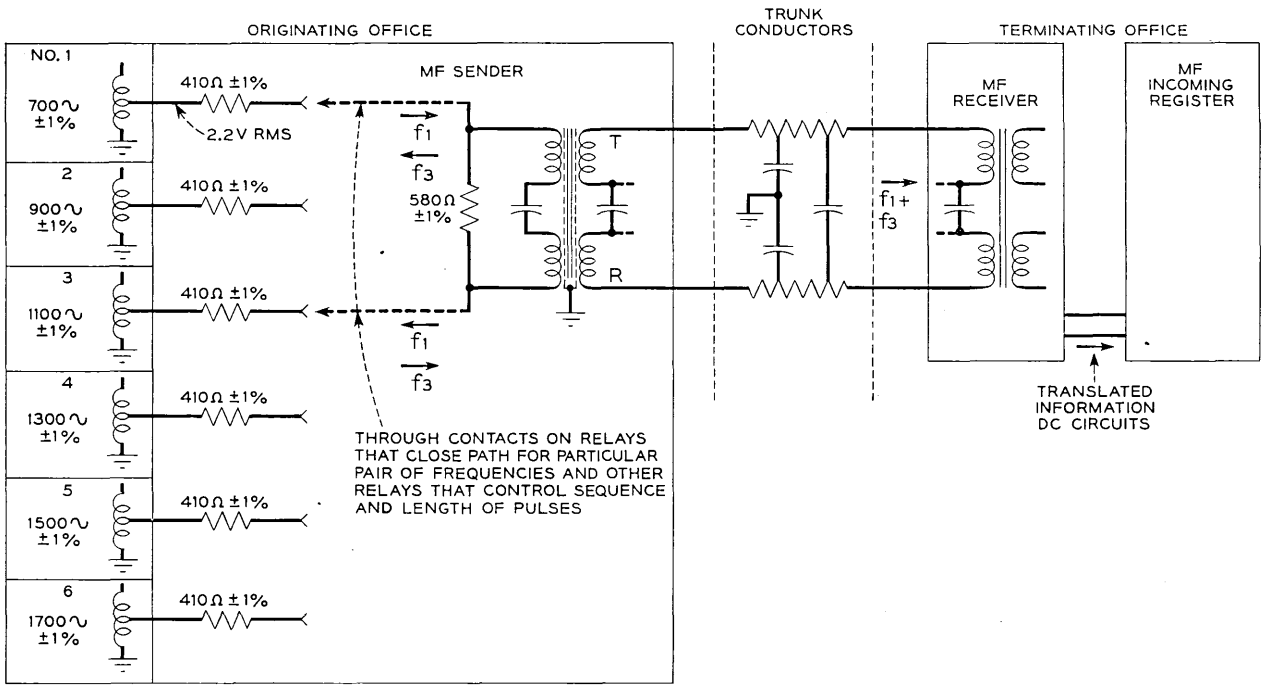


Fig. 1 — Simplified diagram of MF pulsing circuits.

sender becomes very large. It was for this reason that transistor oscillators were developed for use as individual current supplies.

An individual sender will handle a large number of calls during a day but the aggregate time during which the multifrequency current is required is normally less than two hours. The ability of transistors to operate the instant that power is applied is therefore a distinct advantage in saving power. The low voltage that is required for their operation and the long life expectancy are further advantages.

The method used is illustrated in Fig. 1. To transmit a pulse the output windings of two oscillators are connected in series through two 410-ohm resistors and the parallel combination of a 580-ohm resistor and windings on a transformer. The other windings on the transformer are connected to the trunk conductors. At the distant office the trunk is terminated by a similar transformer. The output of this transformer is connected to an amplifier, filter circuits and detectors in an MF receiver. Relays associated with the detectors convert the ac signal into a dc signal for the register circuit.

The 580-ohm resistor across the transformer winding serves as a termination for the trunk between pulses. This resistor in combination with the two 410-ohm resistors in the oscillator circuits serve as the trunk termination during pulsing.

REQUIREMENTS FOR OSCILLATOR

This oscillator was designed to be used in place of an existing piece of equipment. It therefore must have the properties of the existing equipment on which the design of equipments that are associated with it have been predicated.

The requirements are then:

1. Operate at 700, 900, 1100, 1300, 1500 or 1700 cycles per second $\pm$ 1 per cent.
2. Furnish ac signals at a level of -3 dbm $\pm$ 1 db at the input to the trunk.
3. The level difference at all frequencies must be no greater than 1 db.
4. Attain normal output level between the time the sender is seized and the time the first pulse is sent. This time is approximately one half second.
5. Meet all requirements with dc input voltage limits of 45–50 volts.
6. Meet all requirements with ambient temperatures of $+40$°F to $+135$°F.
7. Have level of second harmonic a minimum of 35 db below the fundamental.

8. Provide an output impedance sufficiently low that the differences in impedance of the associated oscillator will not appreciably affect either output level or frequency.

9. Meet frequency requirements when working into trunks that have impedances with phase angles of $+15°$ to $-45°$.

10. Be unaffected by transient disturbances on trunk conductors coupled to the oscillator. These disturbances are caused by lightinng or troubles on closely associated power lines.

DISCUSSION

There are many ways of making a transistor oscillator. One of the simplest and most stable of these uses an LC circuit for controlling the frequency. Additional windings inductively coupled to the tuned circuit inductance provides the necessary feedback to the transistor and the connection to the load. This is the type of circuit adopted for this application. A schematic of the circuit is shown in Fig. 2. The transistor controls power supplied to the oscillatory circuit during only one half the operating cycle. The amplitude of the ac voltage produced across winding 1–2 is almost equal to the dc voltage applied to the collector.

The transistor acts as an amplifier or it may be considered as a negative resistance. Control of the transistor is obtained by current flowing in the emitter that is in phase with the collector voltage. To build up oscillations, the power put into the circuit must be greater than that expended. Stability is obtained when the power put in equals the power expended.

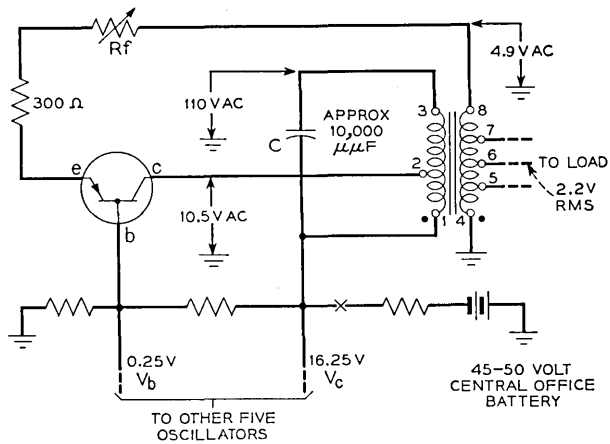


Fig. 2 — Oscillator circuit for MF signaling showing approximate values for voltages.

In the following sections the functions of a resonant circuit are first discussed. Then, the characteristics of a point contact transistor that make it suitable for use in an oscillator are considered. This is followed by a discussion of the behavior of the two when combined to form an oscillator with consideration given to the effect various factors have on performance.

RESONANT CIRCUIT

A resonant circuit has several properties of interest when one considers it as part of an oscillator. These are:

1. The input impedance of a parallel circuit consisting of an inductance, L, and a capacitance, C, has the property of a resistance at approximately frequency,[1]

$$f_0 = \omega_0/2\pi.$$

That is, input impedance,

$$Z = RQ^2\left(1 - j\frac{1}{Q}\right),$$

where R = effective resistance of the inductor and

$$Q = \frac{\omega_0 L}{R}$$

when Q is high, $Z \doteq RQ^2 = R_0$ the equivalent resistance at resonance.

2. The decrement of a parallel resonant circuit when disconnected from sources of power or loads is:[2]

$$\left[\begin{matrix}\text{numerical decrement in cur-}\\ \text{rent or voltage per cycle}\end{matrix}\right] = 1 - e^{-\frac{R}{2Lf_0}}. \tag{1}$$

When the effective resistance R becomes zero the decrement likewise becomes zero.

In the oscillatory circuit considered the decrement is made zero by the use of a negative resistance equal to the effective positive resistance of the turned circuit. Additional loads on the turned circuit are accommodated in the same way.

If the negative resistance in the simple circuit considered becomes greater than the equivalent resistance of the tuned circuit and its loads, power will be supplied by the tuned circuit at a rate that will maintain equality between the power put into the circuit and the power taken out. As a corollary to this, if the negative resistance is less than the positive resistance, the voltage will increase across the circuit. For

the oscillator this would be:

$$\left[\begin{matrix}\text{numerical increment in cur-}\\ \text{rent or voltage per cycle}\end{matrix}\right] = e^{\pi/Q_1} - 1. \tag{2}$$

The subscript on Q is used to denote the condition where R is negative. The equation is obtained from the following relationships.

$$\begin{aligned} Q &= \frac{\omega_0 L}{R} = \frac{L}{R/\omega_0} = \frac{1/2I_m^2 L}{1/2I_m^2 R/\omega_0} \\ &= \frac{\text{Energy stored in resonant circuit}}{\text{Rate of change of energy stored}} . \end{aligned} \tag{3}$$

Since $R = \dfrac{\omega_0 L}{Q}$, a negative R, $(-R) = -\dfrac{\omega_0 L}{Q_1}$,

$$e^{\frac{R}{2Lf_0}} \text{ (from equation 1)} = e^{\frac{-(-R)}{2Lf_0}} = e^{\frac{-(-2\pi f_0 L)}{(2Lf_0 Q_1)}} = e^{\frac{\pi}{Q_1}}. \tag{4}$$

I_m = peak value of current flowing in resonant circuit.

The relationship of energy stored to rate of change of energy stored is a convenient relationship for computing the value of Q_1 in this particular case.

The ability of a transistor to satisfy the requirements of the resonant circuit for sustaining oscillation as well as building oscillation will be discussed later.

Energy which is large in comparison to the power taken by the load is stored in the tuned circuit in order to minimize:

(a) The effect on oscillator frequency of small changes in load current phase angle with trunks of different types, connected.

(b) The effect on oscillator frequency of different values of inductance in the output windings of the associated oscillator circuit.

(c) The effect of surges on the transistor that are transmitted over the connected trunk to the oscillator.

(d) The harmonics produced by the method of supplying power to the tuned circuit. This includes the use of an incorrect feedback adjustment.

The Q for the transformers varied from 60 to 85. The stored energy was then 60 to 85 times the energy dissipated per radian at the oscillaing frequency for the particular transformer. By adjustment of the ratio of the turns between windings 1–2 and 1–3 a dissipation in the tuned circuits of from 6 to 10 milliwatts was obtained.

In the six-frequency supply three different coil designs are used. One design is used for the two lower frequencies, a second for the two inter-

mediate frequencies and a third for the last two frequencies. Data on these coils are given in Appendix I.

The frequency of oscillation for each coil is determined by the capacitance used with it and is given approximately by the formula:

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{1}{LC}}.$$

Sufficient inductance is used in these coils so that mica condensers can be economically used with them. Mica condensers are used because of their low temperature coefficient.

Leakage reactance and capacitance between windings make the circuit resonant at more than one frequency. To minimize the current fed back to the emitter under such parasitic conditions, the 4–8 winding was placed next to the core with terminal 8 next to terminal 1. This placed an ac ground next to the end of the winding connected to the emitter. As a further precaution against parasitic oscillations the resistance between emitter and transformer is kept high.

TRANSISTOR OPERATION

It is customary to consider the transistor as an amplifier working into a load represented by the tuned circuit. However since the current in the collector and that in the emitter are intimately related during that part of the cycle when power gain is obtained, the collector circuit can be considered equally well as a negative resistance of a value established by the feedback used and the emitter circuit simply as a load. At best either method is only an approximation due to the nonlinearity of the transistor characteristics for large signals. Representation of the transistor as a resistance puts the requirements in terms of values readily obtained from the static characteristics of a transistor. This form of treatment is therefore used.

The regions in which positive or negative resistance is obtained is illustrated in Fig. 3. The characteristics shown are those for an ideal transistor. That is, the ratio of an incremental change in collector current to an incremental change in emitter current, α, with a constant collector voltage is a uniform value in region 2. Also, that in regions 1 and 3 the slope of the lines in each region is constant. That this is not very different from that obtained from some transistors can be seen by comparison with the actual characteristics shown in Fig. 4.

The division line between regions 1 and 2 represents the magnitude of the dc voltage applied between collector and base, V_c-V_b. On the left hand side of Fig. 3 the phase relationship and magnitude of ac voltage

applied to the collector and the emitter current are shown. The condition which limit the ac voltage on the collector to this value will be discussed later.

In region 1 the net voltage on the emitter to base circuit is negative and it has no effect on the current which flows in the collector. The current which flows in the collector will therefore depend purely on the reverse resistance of the collector acting as a diode and the voltage applied between collector and base. This is a positive resistance but it is not always linear. Since the only voltage active in this region is that obtained from the tank circuit, this resistance acts as a load on the tank circuit.

In region 2 the voltage applied to the emitter circuit is positive and the current that flows as a result of this voltage exercises control over the collector current. The external resistance that is used in the emitter circuit is sufficiently high so that the small changes that occur in the emitter input resistance as the emitter current is varied are insignificant. The current that flows in the emitter circuit can therefore be considered vary linearly with the ac voltage. Since for each incremental change in collector voltage a proportional change in emitter current will be obtained, a plot of this relationship for a given ratio of emitter to collector ac voltage will result in a straight line. The slope will be determined by the amount of feedback (emitter voltage). Since a decrease in the absolute voltage on the collector results in an increase in collector current, the line has a negative slope. It therefore represents a negative resistance.

In region 3 the emitter current exercises very little control over the collector current.

A negative resistance represented by $-R2$ is therefore obtained from 0 to π of the ac wave. If a second transistor were added of identical charac-

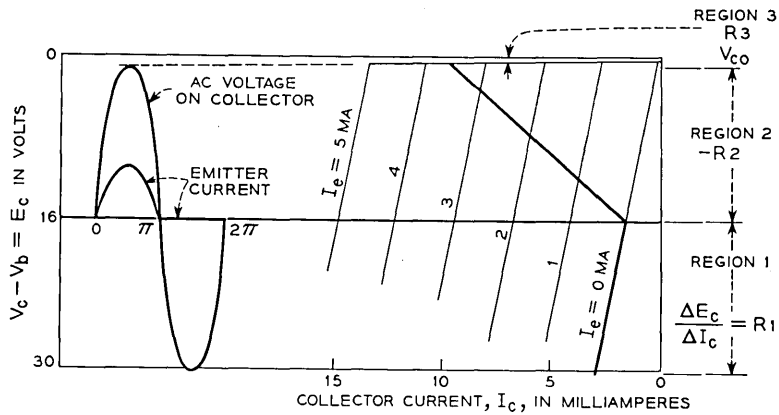


Fig. 3. — Idealized transistor characteristics with operating regions for oscillator indicated.

teristics so connected (push-pull circuit) that a negative resistance would be obtained from π to 2π of the ac wave, the effect of a negative resistance of constant value would be obtained. In order to reduce the transistors to the terms of a two-pole device having a negative resistance the positive resistances represented by the emitter circuits and regions 1 of the transistors would also have to be included.

It is more convenient however to combine all elements that produce a loss when using the static characteristics of a transistor to determine if the transistor will satisfy the circuit requirements for oscillation. This method is therefore used when considering the circuit operation.

There is one important difference between the characteristics of a transistor and the ideal characteristics shown. That is, for very small values of emitter current bias and a constant E_c, the ratio $\Delta Ic/\Delta Ie$, or, α, drops rapidly from three or more to approximately one as the emitter current approaches zero. To eliminate this change in the dynamic negative resistance for very low voltage changes on the collector, a small dc current is supplied in the emitter circuit. The need for this will be discussed later.

CIRCUIT OPERATION

The negative resistance of the transistor is effectively connected in parallel with the resistances representing the various elements that make up the total loss by transformer action. The requirements for oscillation are therefore met when $-R2$ for a complete cycle is less than the positive resistance representing all losses. If the length of time that $-$R2 is effective is only one half cycle the value of $-R2$ must of course be cut in half.* Since a single transistor could meet the requirement for $-R2$ only one was used in the working circuit as is shown in Fig. 2.

The transistor adopted for this use was the 1729 type, now in production, which has been given the RTMA designation 2N25. The 1729 type was used because its characteristics are least affected by changes in temperature, and in addition the allowable power dissipation was approximately twice that of other comparable transistors.

The various factors that when combined make up the load and their normal variation are given in Table I. All losses are in terms of power into the 1–2 winding.

The corresponding value of load resistance is,

$$R_{TL} = \frac{(V_c - V_b - V_{co})^2 1000}{2 \times 2 \times W_t}.$$

* See Appendix II.

TABLE I

Source of Loss	Power in Milliwatts		
	Min.	Avg.	Max.
Sustaining stored energy in resonant circuit	6.0	8.0	10.0
Output load	4.5	4.5	4.5
Region 1 of collector operation	2.8	7.0	14.0
Loss in emitter	4.0	8.5	13.4
Margin for stability of adjustment	12.0	12.0	12.0
Total, W_t	29.3	40.0	53.9

This would be 1920-ohms, 1400-ohms and 1045-ohms respectively. For purpose of illustration the average value is plotted on the characteristics of an average transistor in Fig. 4 along with $-R2$ for the transistor. It is evident from this that $-R2$ is lower in value hence the circuit will oscillate and build up to the required voltage. Minor corrections in the emitter current would normally be required in order to meet test requirements. The potentiometer that is shown in Fig. 2 provides the means for adjustment.

The oscillogram shown in Fig. 5 illustrates the condition described above. The characteristics of the transistor used are shown in Fig. 4. The oscillogram is a multiple exposure from which $R1$, $-R2$ and $R3$ (see Fig. 3) for several values of feedback may be obtained. The normal condition of adjustment is illustrated in Fig. 6 with normal load. Four times normal load is a test condition.

When the extra load that is applied during test is removed the output voltage should go up since power input exceeds the power expended.

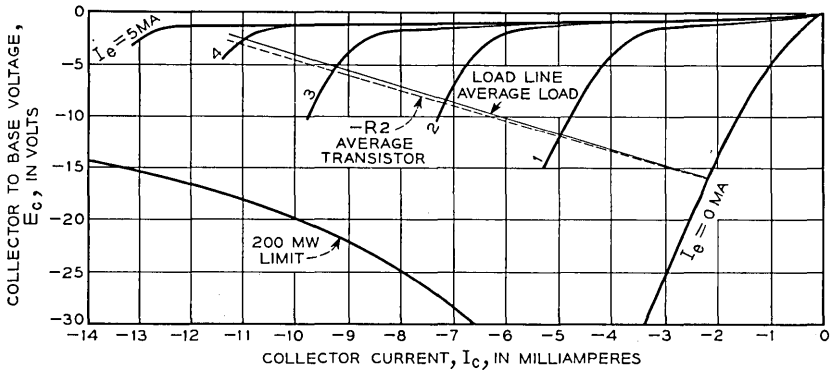
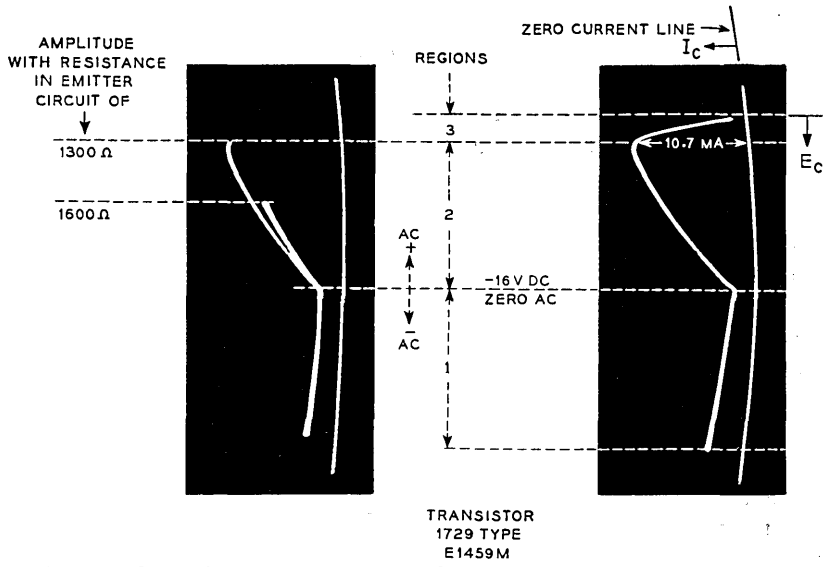


Fig. 4 — Characteristics of representative 1729 type transistor with negative resistance values plotted for the average condition given in table. A load line is also shown for the average condition. In plotting $-R2$ a line is drawn from $I_e = 0$, $E_c = 16$ to point corresponding to maximum I_e. Maximum $I_e \doteq W_e \times \frac{1.414 \times 2}{E_{erms}}$ at $E_c = 1.5$ volts. W_e is loss in emitter circuit given in table.



Figs. 5 and 6 — Oscillograms showing the relationship of collector current to collector voltage during a complete cycle for several operating conditions. Fig. 5 (left) is a multiple exposure made to illustrate the ability to oscillate at a lower output level with decreased feedback. This is due to the much higher alpha obtained with low emitter currents. Fig. 6 is for normal operating conditions.

This increase in voltage is however very small since the losses will increase as the square of the voltage and the rate at which energy is supplied to the circuit will decrease. The decrease in rate is due to the change from a negative resistance to a positive resistance in region 3. This causes the average negative resistance for the complete cycle to have a higher value.

Feedback which is far greater than is required can in some cases cause the peak value of the ac voltage to exceed the dc voltage. Power is drawn from the energy stored in the tank circuit when this occurs. This effectively limits any further increase in output voltage.

A change will also be introduced in the emitter circuit due to operation in region 3. That is, the voltage feedback introduced in the emitter circuit by collector current flowing in the common base resistance is reversed in phase in region 3.[3] This is due to a reduction in collector current when the voltage applied to the emitter circuit is still rising. This feedback is sufficient in many cases to cancel the increase in emitter voltage. The emitter current in such a transistor will therefore remain nearly constant in this region. In region 2 however the feedback is in such a direction as to aid the flow of emitter current. The result is that the voltage drop across the emitter resistance is approximately canceled by the voltage across the base resistance. Due to this relationship the tran-

sistor may appear to have zero resistance in the emitter to base circuit or a small positive or negative resistance.

It is evident from the foregoing that the ability of a transistor to fulfill the requirements for oscillation is therefore dependent upon both the average α along the load line and, $R1$, resistance in region 1. The relationship of these two factors is shown in Fig. 7. Dots on this chart represent the 1729 type transistors tested that met the requirements for the 2N25 transistor. The ambient temperature was +135°F. It is inadvisable to use transistors having a resistance of much less than 4000-ohms in region 1 since both the average dissipation rate and the peak dissipation rate would exceed allowable limits for continuous operation. This will tend to cause the transistor characteristics to change at a more rapid rate.

The effect of different values of feedback on the output of the oscillator is shown in Fig. 8. Fig. 8 also shows the variation in output obtained with several different values of load resistance. This was done to illustrate the use of increased load in determining the proper point for adjusting the feedback resistance. The proper adjustment is the minimum feedback with which the output changes only approximately 1 db in going from normal load to four times normal load. This degree of margin

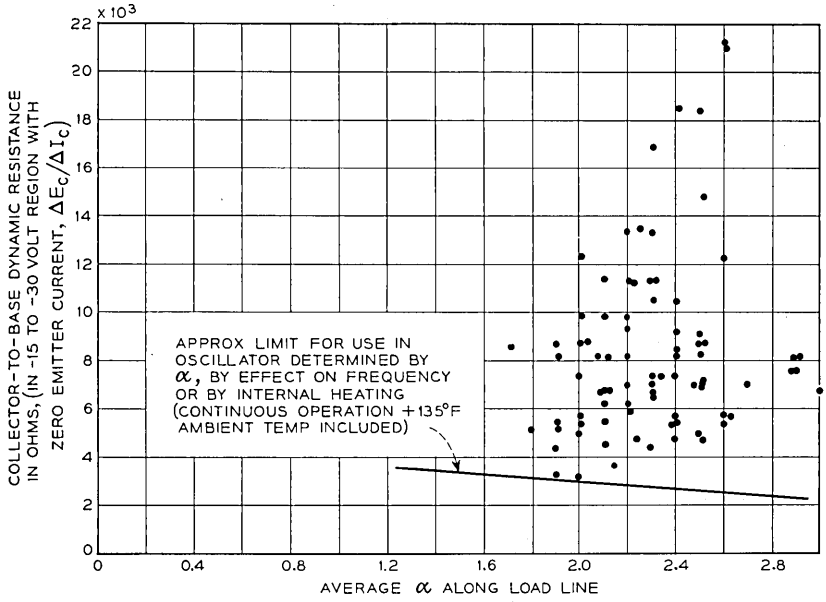


Fig. 7 — Plot of α versus collector to base resistance for representative group of transistors meeting 2N25 transistor requirements.

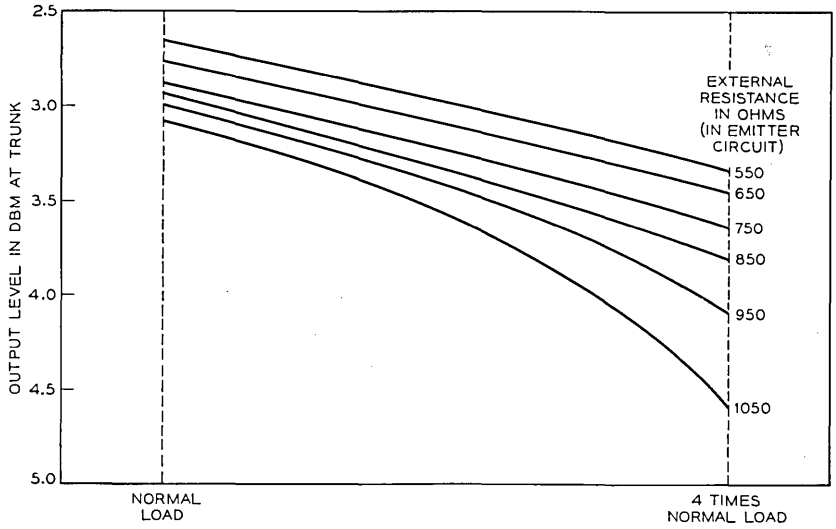


Fig. 8 — Effect of load and value of feedback resistance on the level of the output.

permits some deterioration in transistor characteristics before the change in output is sufficient to require readjustments.

Variation in the absolute level of output due to variations in V_{co} (see Fig. 3) between transistors and to differences in coils is taken care of by the use of taps on the output windings.

Harmonics of the fundamental frequency are created by the nonlinearity of the transistor characteristics. These harmonics are accentuated by excessive feedback. The level of the harmonics for a representative transistor are shown in Fig. 9.

The effect of variations in feedback on the frequency is shown in Fig. 10. The shift is thought to be due to several factors all small. One is the lack of perfect coupling between the transformer windings. Another is

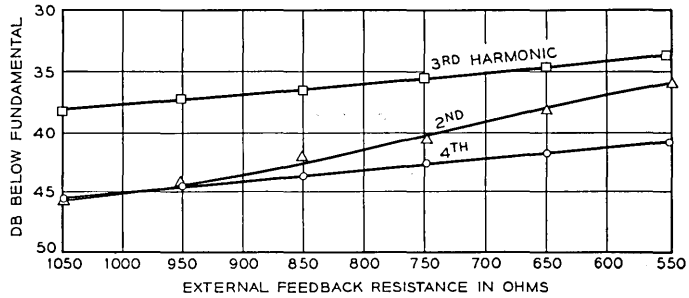


Fig. 9 — Effect of feedback on harmonics with normal load.

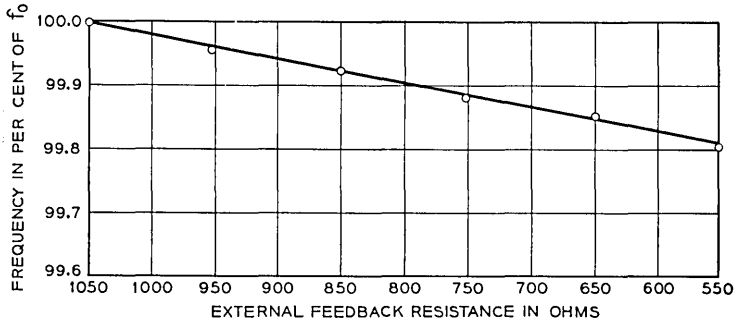


Fig. 10 — Effect of feedback on frequency.

the leakage reactance of the input winding. As the time rate of change in current is increased by increased feedback these factors become increasingly greater although never very large. Fewer turns are required on the coils used for operation at the higher frequencies hence these effects are reduced.

The output level of the oscillator will vary almost directly with the variations in the dc voltage since the amplitude of ac voltage across collector to base is almost equal to the dc voltage applied. Hence, a variation of approximately 0.9 db will be obtained in the output when the central office battery is reduced from 50 volts to 45 volts due to power failure conditions.

The over-all output variation from all causes is shown in Fig. 11. This is based on data obtained using the transistors having the distruibtion in characteristics shown in Fig. 7.

Decreases in the value of R1 with temperature is normally compensated by a corresponding increase in α. However small positive or negative voltage changes that alter the level of output do occur in the cut-off voltage. This is minimized by keeping the dc voltage as high as permissible and still meet the 200 milliwatt dissipation limit for the 2N25 transistor.

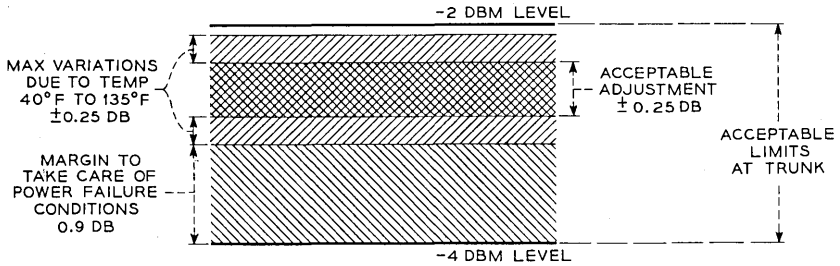


Fig. 11 — Output level with effect of various factors that may alter the level indicated.

The effect of the various factors mentioned before on the frequency of operation are shown in Fig. 12. Since several of the factors causing a shift in frequency were in the negative direction only, the adjustment limits were set correspondingly higher. The over-all frequency variations could be reduced by reducing adjustment tolerances.

The starting condition is important in this type of circuit since energy must be introduced into the oscillatory circuit before the dynamic characteristics of the circuit become effective. This means that the build up time is dependent upon the amount of energy introduced into the system at the start. In this application energy is introduced by the current which flows when the dc voltage is applied to the collector circuit. The value of this current is largely dependent upon the collector to base dc resistance

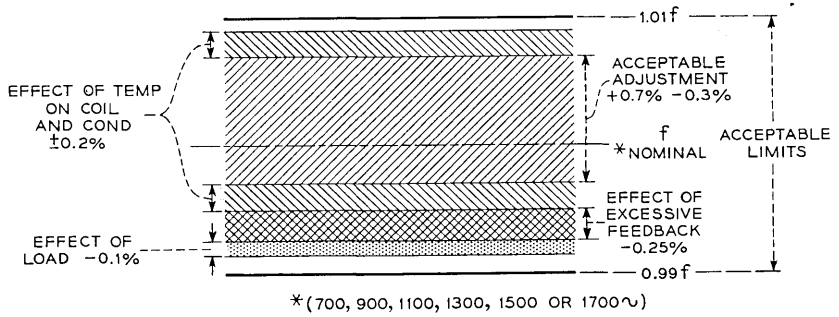


Fig. 12 — Frequency of output with effect of various factors that may alter the frequency indicated.

which is in turn affected by the ambient temperature. This resistance is between approximately 4,000 and 20,000 ohms.

The oscillograms shown in Fig. 13 illustrate this effect. Both are for the same circuit operating under normal conditions of adjustment. Q_1 for this condition is approximately 18. Oscillogram (a) is for the application of voltage in the normal fashion to the voltage divider circuit. The closure occurs at the point oscillation starts. Oscillogram (b) shows the build up obtained when no impulse is applied to start oscillation except for minor irregularities in the dc voltage applied. Starting is prevented in this case by a short circuit on 4–5 widing that was removed approximately 2 ms after the start of the trace. The amplitude of oscillation is so low however for the first few cycles that the start is difficult to distinguish. The build up time for (a) is approximately 27 milliseconds and for (b) it is approximately 37 milliseconds. The exponential build up of amplitude is modified greatly by the rapid change in the transistor's α as the voltage approaches the cutoff region.

It should be noted also that oscillation would not have started under

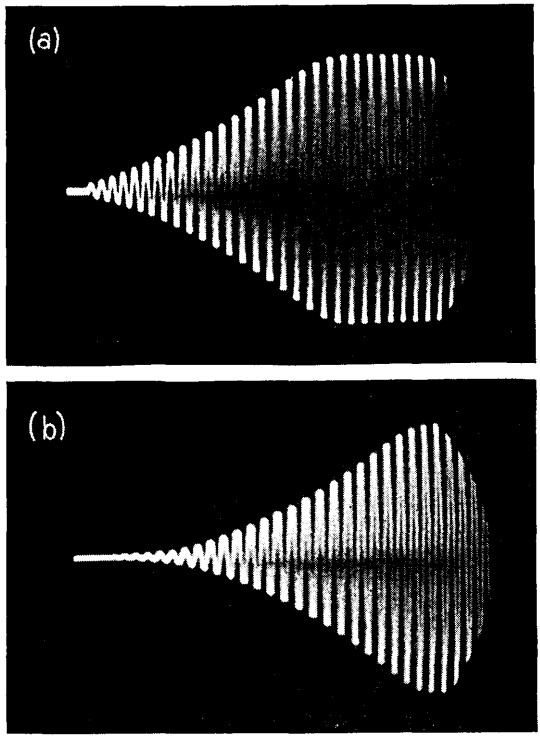


Fig. 13 — Build-up of oscillation in resonant circuit. Normal operating conditions.

the condition for oscillogram (b) with some transistors if the small emitter bias current had not been provided.

TRANSIENT EFFECTS

The trunk conductors are balanced with respect to ground. Voltages set up in these conductors due to electrostatic or magnetic coupling to the source of the interference will cause longitudinal currents to flow. An electrostatic shield in the transformer, shown in Fig. 1, effectively prevents such longitudinal currents from reaching the oscillator circuit.

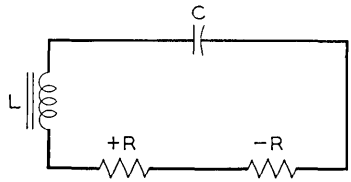


Fig. 14 — Equivalent resonant circuit.

If however the voltage becomes sufficient to breakdown one of the protector blocks that are connected between each trunk conductor and ground, a voltage comparable to the breakdown potential of the protector blocks (400 to 600 volts) would then be impressed across the output windings of the transformer. The usual cause of this condition is lightning. However, neither artificially simulated lightning nor transients of longer duration were capable of raising the voltage on the tank circuit to the point where a transistor was damaged. This is due both to

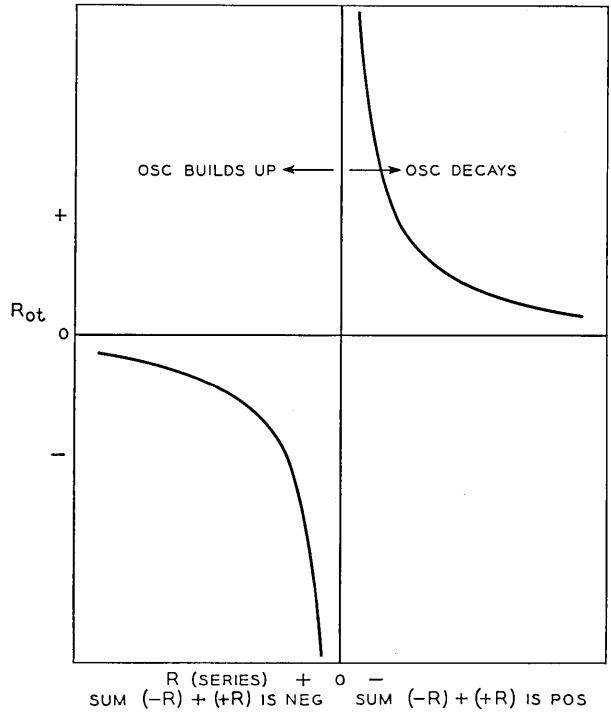


Fig. 15 — Relationship of shunt to series resistance.

the high energy level of the tank circuit and the isolation furnished between repeating coil and oscillator by the series resistors.

SUMMARY:

The transistor oscillator adequately fulfills the requirements for a source of current for multifrequency pulsing over telephone transmission circuits. Adjustments are provided so that the requirements for frequency stability, harmonic level and output level can be met with transistors having a wide range of characteristics. Sufficient margin is provided by

design and by initial adjustment so that an appreciable change in transistor characteristics can be tolerated before readjustment is required.

ACKNOWLEDGMENT

The transformer used in the oscillator was designed by H. E. Vaiden and A. M. King. The transistor used was developed by R. J. Kircher and N. J. Herbert. D. J. Houck assisted in testing the circuit.

REFERENCES

1. Lepage Seely, *General Network Analysis*, McGraw Hill.
2. Kurtz and Corcoran, *Introduction to Electric Transients*, John Wiley and Sons.
3. R. M. Ryder and R. J. Kircher, Some Circuit Aspects of the Transistor, B.S.T.J., **28**, pp. 367–401, July, 1949.

APPENDIX I

Transformers moly-permalloy dust core 1.57 O.D.

700- and 900-cycle operation

Winding (1–2) = 500 turns — Q at 700-cycles, 60
Winding (4–5) = 107 turns — Q at 900-cycles, 68
Winding (4–6) = 115 turns
Winding (4–7) = 123 turns
Winding (4–8) = 220 turns
Winding (1–3) = approximately 5560 (adjusted to meet inductance requirements of 5.2H ± 1 per cent)

1100- and 1300-cycle operation

Winding (1–2) = 372 turns — Q at 1100-cycles, 73
Winding (4–5) = 80 turns — Q at 1300-cycles, 75
Winding (4–6) = 86 turns
Winding (4–7) = 92 turns
Winding (4–8) = 164 turns
Winding (1–3) = approximately 4080 (adjusted to meet inductance requirements of 2.8H ± 1 per cent)

1500- and 1700-cycle operation

Winding (1–2) = 305 turns — Q at 1500-cycles, 82
Winding (4–5) = 66 turns — Q at 1700-cycles, 85
Winding (4–6) = 71 turns
Winding (4–7) = 75 turns
Winding (4–8) = 135 turns
Winding (1–3) = approximately 3360 (adjusted to meet inductance requirements of 1.9H ± 1 per cent)

Appendix II

The equivalent circuit for the oscillator is shown in Fig. 14. For stable operation from equation (1), since $f_0 = 1/t$,

$$1 - e^{-\frac{(+R)\,t}{2L}} = e^{-\frac{(-R)\,t}{2L}} - 1.$$

The negative resistance, $(-R)$, must therefore be equal to the positive resistance $(+R)$. If however $(-R)$ is active for only half the time, $(-R)$ must be equal in magnitude to $2(+R)$ in order to satisfy the requirements for equality. This assumes that boundary effects are negligible. This assumption was borne out by experiment.

The equivalent resistance of the resonant circuit is

$$R_0 \doteq \left(\frac{\omega_0 L}{R}\right)^2 R \doteq K\frac{1}{R},$$

where

$$k = (\omega_0 L)^2.$$

A plot of this relationship when positive and negative resistances are combined, is shown in Fig. 15.

In the actual circuit the negative resistance is connected across the 1–2 winding. The 1–3 winding is in the resonant circuit. The equivalent resistance ($R_{0(1\text{-}2)}$ of the resonant circuit across winding 1–2, is determined as follows:

$$R_{0(1\text{-}2)} = \frac{(\text{turns, winding } 1\text{–}2)^2}{(\text{turns, winding } 1\text{–}3)^2} \times R_0.$$

Ferrites in Microwave Applications

By J. H. ROWEN

(Manuscript received May 26, 1953)

Since Hogan's exposition of the extreme usefulness of the microwave Faraday effect numerous other laboratories have begun investigating propagation through ferrites and have made significant contributions to the art. In view of the tremendous interest which is being accorded this work this paper has been prepared to summarize some of the observations and developments to date. The plane wave theory is reviewed briefly with special attention being given to the mechanisms by which power is absorbed by the ferrite. The plane wave theory is then modified to describe various waveguide effects. Finally experimental procedures and results are presented to illustrate the theory and to provide general information regarding the design of devices employing these effects.*

INTRODUCTION

The ferromagnetic Faraday effect occurs at microwave frequencies as a direct result of the dispersion in permeability which is associated with ferromagnetic resonance. The resonance can be explained most simply by stating that the total magnetization vector of a magnetized ferromagnetic material has associated with it an angular momentum arising from the angular momenta of all of the spinning electrons contributing to the magnetization. Because of this angular momentum (which is directed along the same axis as is the magnetic moment but in the opposite direction) the magnetization vector behaves as a top or gyroscope. If it is displaced from its equilibrium position in a steady magnetic field it will not rotate directly into alignment with the field but will precess about the dc field direction at a frequency determined by the strength of the dc field. In the absence of damping this precession would continue indefinitely, but damping losses are such that the precession will damp out in approximately 10^{-8} sec.

* C. L. Hogan, The Ferromagnetic Faraday Effect at Microwave Frequencies and Its Applications — The Microwave Gyrator, B.S.T.J., **31**, pp. 1–31, Jan., 1952.

If an alternating field is applied at right angles to the dc field the magnetization will be driven in precession and when the driving frequency coincides with the natural resonance frequency as determined by the strength of the dc field a large amount of power will be absorbed from the driving field. Off resonance the power absorption is small, but the effective permeability seen by the driving field will go through a dispersion such as is exhibited by all resonant systems. With this model in mind we can proceed to discuss the phenomenon of ferromagnetic resonance and the ferromagnetic Faraday effect.

INFINITE MEDIUM — LONGITUDINAL FIELD

Polder[1] has shown that because of the gyroscopic nature of the magnetization a tensor permeability is required to relate the magnetic flux density and field intensity vectors in a ferromagnetic medium. At low frequencies the off-diagonal components of this tensor are negligible, and the tensor reduces to the ordinary scalar permeability. At frequencies above about 100 mc. these off-diagonal components can become significant depending upon the magnetic state of the material. When this tensor permeability is introduced into Maxwell's equations and a wave equation is derived for propagation in the direction of the applied magnetic field we find that the normal mode solutions to the wave equation are two circularly polarized waves rotating in opposite directions. A solution in terms of linear polarizations is, of course, possible; but the result is more readily interpreted in terms of the circularly polarized waves. Furthermore, the propagation constant for either circular wave contains a simple scalar permeability, instead of the tensor required to describe the medium in general.

Polder's tensor permeability gives the following relations between $\mathbf{b}$ and $\mathbf{h}$ when there is a static magnetic field along the positive z axis*.

$$\begin{aligned} b_x &= \mu_0\mu h_x - j\mu_0\kappa h_y \\ b_y &= j\mu_0\kappa h_x + \mu_0\mu h_y \\ b_z &= \mu_0 h_z \end{aligned} \tag{1}$$

The quantities

$$\mu = \mu' - j\mu'' \tag{2}$$

$$\kappa = \kappa' - j\kappa'' \tag{3}$$

are complex relative diagonal and off-diagonal components of the tensor permeability.

[1] D. Polder, Philosophical Mag., **40**, p. 99, Jan. 1949.

* Lower case letters are used for RF magnetic quantities.

Equations giving μ and κ in terms of the applied magnetic field and the fundamental atomic constants are given by Hogan.[2] These equations show that both μ and κ have a resonance at a value of effective internal field given by $\gamma H = 2\pi f$. If H is in ampere turns per meter and f in megacycles the gyromagnetic ratio is $\gamma/2\pi = 3.51 \times 10^{-2}$.

A plane wave travelling in the z direction in an infinite medium described by equations (1) can be resolved into two counter-rotating circularly polarized plane waves having propagation constants as follows:

$$\Gamma_+ = j\omega\sqrt{\mu_0\epsilon_0}\,\sqrt{\epsilon(\mu - \kappa)} \qquad \Gamma_- = j\omega\sqrt{\mu_0\epsilon_0}\,\sqrt{\epsilon(\mu + \kappa)} \qquad (4)$$

where the subscripts $\pm$ refer to positive and negative circularly polarized waves.* The terms inside parentheses are effective scalar permeabilities completely describing the medium for a circularly polarized wave. Calculated values of these effective permeabilities are plotted as functions of applied field in Fig. 1 for a ferrite operating at three frequencies.

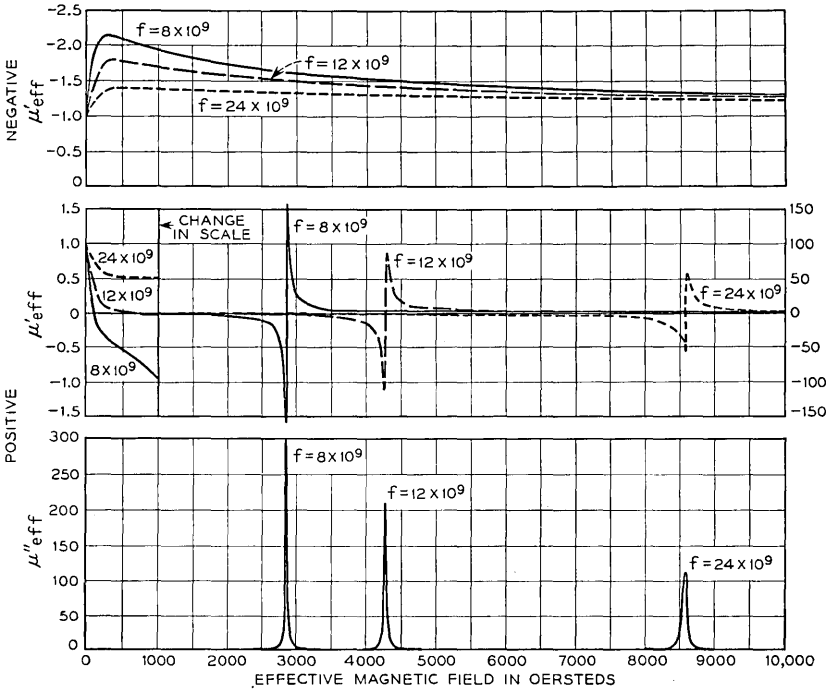


FIG. 1 — Calculated effective permeabilities for positive and negative circularly polarized plane waves computed for three different wave frequencies.

[2] C. L. Hogan, Revs. Mod. Phys., **25**, Jan., 1953.

* A positive circularly polarized wave is one which rotates in the direction of the positive current producing the dc magnetic field.

From these curves we may calculate the rotation per unit path length, the absorption of the positive component, the net insertion loss and the ellipticity of the resultant wave.

The rotation per unit length is given by

$$\frac{\theta}{l} = \tfrac{1}{2}(\beta_- - \beta_+) \tag{5}$$

where $\beta_\pm$ are the imaginary parts of the propagation constants, $\Gamma_\pm$. Let us consider the special case in which the dielectric loss is zero. For convenience we define the complex effective permeabilities seen by the circularly polarized waves as follows:

$$\begin{aligned} \mu_+ &= \mu - \kappa = \mu_+' - j\mu_+'' \\ \mu_- &= \mu + \kappa = \mu_-' - j\mu_-'' \end{aligned} \tag{6}$$

The propagation constants may then be written:

$$\begin{aligned} \Gamma_+ &= \omega\sqrt{\mu_0\epsilon_0}\sqrt{\frac{\epsilon}{2}}\left[\sqrt{|\mu_+| - \mu_+'} + j\sqrt{|\mu_+| + \mu_+'}\right] \\ \Gamma_- &= \omega\sqrt{\mu_0\epsilon_0}\sqrt{\frac{\epsilon}{2}}\left[\sqrt{|\mu_-| - \mu_-'} - j\sqrt{|\mu_-| + \mu_-'}\right]. \end{aligned} \tag{7}$$

It is of particular interest to consider what happens to β_+ when μ_+ becomes zero or negative. If we rewrite the expression for β_+:

$$\beta_+ = \frac{\omega}{c}\sqrt{\frac{\epsilon}{2}}\sqrt{\sqrt{(\mu_+'^2 + \mu_+''^2)} + \mu_+'} \tag{8}$$

we see that, when μ_+' is zero or negative, β_+ depends primarily on the magnitude of μ'', for wherever μ'' is negligible, β_+ is zero. Furthermore, we see that the attenuation constant, α_+, given by:

$$\alpha_+ = \frac{\omega}{c}\sqrt{\frac{\epsilon}{z}}\sqrt{\sqrt{\mu_+'^2 + \mu_+''^2} - \mu_+'} \tag{9}$$

becomes dependent primarily upon μ_+' when μ_+' becomes negative so that we observe a significant attenuation long before μ_+'' becomes large. In Fig. 2 are shown the rotation of the plane of polarization of the linearly polarized wave and the absorption of the positive circularly polarized component of the wave. The dielectric constant of the ferrite was assumed to be 9.0, a typical value for many ferrites.

From these curves it is evident that the wave will be elliptically polarized whenever the effective field is large enough to make μ_+' zero or

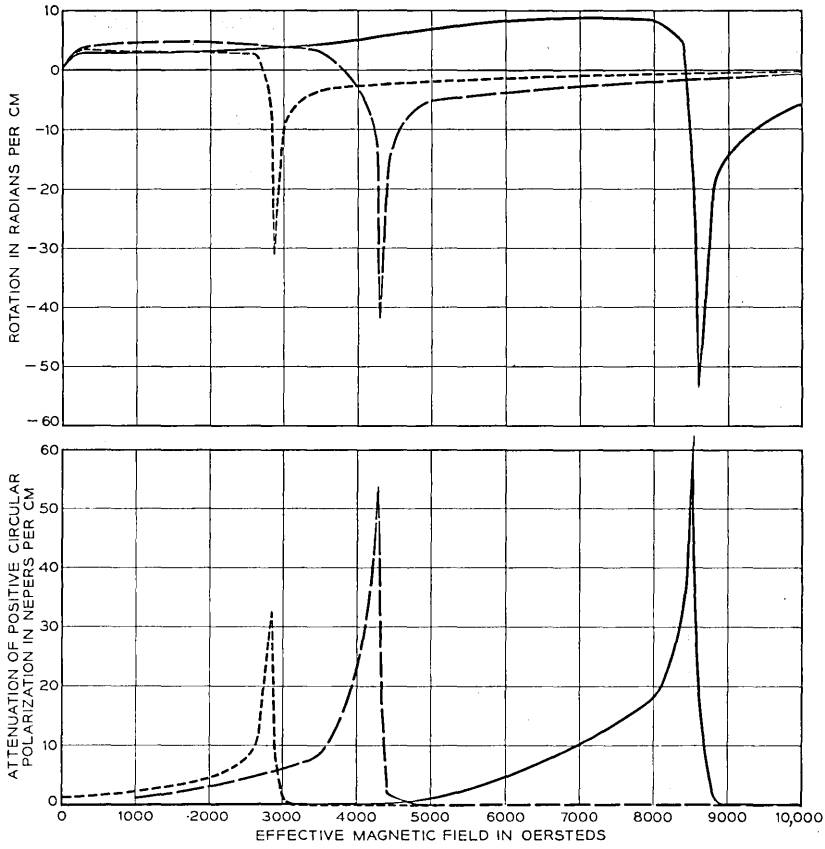


FIG. 2 — Rotation of the linearly polarized plane wave and absorption of positive circularly polarized component versus effective static field computed from data of Fig. 1.

negative, the amount of ellipticity depending in part upon the distance the wave has traveled in the ferrite medium.

PLANE WAVE, TRANSVERSE FIELD

If a wave is propagated in either the x or y direction when the dc field is in the z direction then the wave equation has two orthogonal solutions representing linearly polarized waves. One of these is polarized with the electric vector parallel to the applied dc field and the other has the magnetic vector parallel to the applied dc field. When the magnetic

vector of the wave is parallel to the magnetization the torque on the electrons is zero and the wave sees an isotropic dielectric medium with relative permeability equal to unity. However, when the electric vector is parallel to the magnetization the magnetic vector is at right angles to it and can set the electrons into precession.

Consider a wave propagating in the x direction in an infinite medium magnetized in the z direction. Let this wave be polarized so that it has components E_z and h_y. When this wave enters the magnetized medium, h_y exerts a torque on the magnetization vector M causing it to precess about the z axis. This results in both an m_y and an m_x component of alternating magnetization. There is, however, no component of $\mathbf{b}$ in the x direction because of internal demagnetizing fields arising from an effective volume distribution of magnetic charge as shown in Fig. 3. Such a volume distribution of magnetic charge arising from the periodic reversal in phase of the driving magnetic field is propagated through the medium at the velocity

$$v = \frac{c}{\sqrt{\mu\epsilon}}$$

An instantaneous picture of this distribution is shown in Fig. 3. The magnetic poles set up a magnetic field in the x direction throughout the material. This field is commonly called a demagnetizing field and for

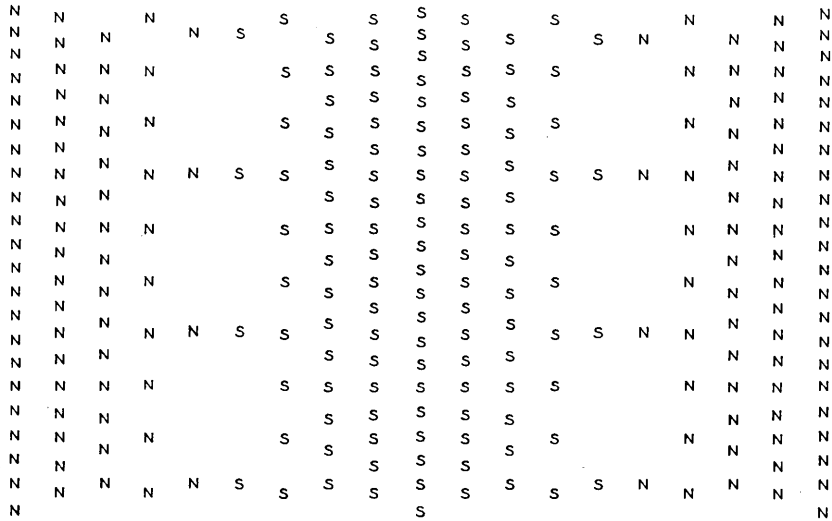


FIG. 3 — Effective volume distribution of magnetic poles arising from phase reversals in transversely magnetized infinite medium.

this particular case it can easily be shown that at every point in the material this field is given by*

$$h_x = -m_x$$

and by definition

$$b_x = \mu_0(h_x + m_x) = 0$$

In one sense this wave is no longer a plane wave as it has a component of **h** in the direction of propagation. However, the electric field, E, and the magnetic flux density, **b**, are unchanged and remain the same as in a normal plane wave.

The solution to the wave equation for the foregoing case yields a propagation constant

$$\Gamma = j\omega\sqrt{\mu_0\epsilon_0}\sqrt{\frac{(\mu^2 - \kappa^2)\epsilon}{\mu}} \tag{10}$$

in which the effective relative permeability of the medium is

$$\mu_{\text{eff}} = \frac{\mu^2 - \kappa^2}{\mu} \tag{11}$$

The real and imaginary parts of this expression are plotted in Fig. 4 for three frequencies.

These curves have the same general shape as those for the *positive* circular component of the wave propagated along the dc field direction, but here they apply to the entire linearly polarized wave. Again we have the possibility of zero or negative permeability. In the region just above resonance the real part of the permeability takes on large values and maintains these even after the absorption curve is nearly zero. This suggests that it is possible to adjust the permeability to equal the dielectric constant of the material so that the medium matches free space perfectly. The medium then can be used as a switch by changing the field from the point where $\mu_{\text{eff}} = 0$ to the point where $\mu_{\text{eff}} = \epsilon$.†

In the region between zero applied field and saturation where the curve levels off, the effective permeability changes almost linearly. In the

* We follow Stratton in making M an H-like quantity rather than B-like.

† In a waveguide μ_{eff} must be adjusted to satisfy the condition that

$$1 = \mu_{\text{eff}}\sqrt{\frac{1 - \left(\frac{\lambda}{\lambda_C}\right)^2}{\mu_{\text{eff}}\epsilon - \left(\frac{\lambda}{\lambda_C}\right)^2}}$$

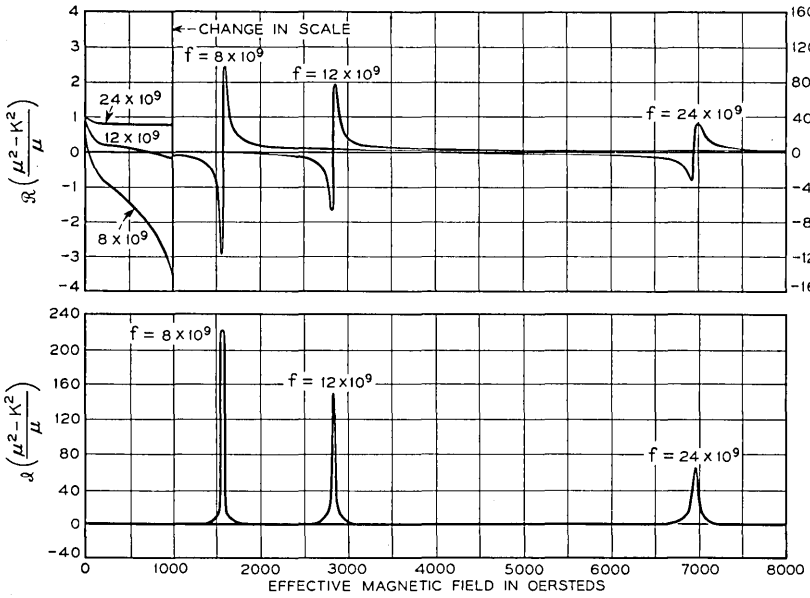


FIG. 4 — Effective permeability seen by a plane wave in a transversely magnetized infinite medium. Real part above and imaginary part below.

absence of dielectric loss the propagation constant of the wave is given by:

$$\beta = \omega\sqrt{\mu_0\epsilon_0}\,\sqrt{\epsilon}\,\sqrt{\frac{(\mu^2 - \kappa^2)}{\mu}} = \omega\sqrt{\mu_0\epsilon_0}\,\sqrt{\epsilon}\,\sqrt{\mu_{\text{eff}}} \qquad (12)$$

Thus a variable phase delay is obtained by controlling the magnitude of the applied field. This delay is not *necessarily* accompanied by a change in attenuation so that an ideal phase shifter can be made using this effect. There are numerous other applications which can be made of the longitudinal and transverse applied field phenomena. Many of these will be discussed later in this article.

LOSS MECHANISMS IN FERRITES AT MICROWAVE FREQUENCIES

Neglecting dielectric losses, the plane wave theory predicts almost no loss at all for a negative circularly polarized wave and a single absorption line for a positive C.P. wave. In practice a more complicated behavior is observed, and to facilitate the discussion we show in Fig. 5 typical loss characteristics superposed upon the theoretical loss curve of Fig. 1. We will enumerate the main points of interest before proceeding with the discussion. The specific differences in behavior are:

Curve A. Broad resonance absorption line.

Curve B. A loss which disappears when the material is magnetized, called "Low Field Loss".

Curve C. Loss which goes to zero for one component and rises for the other.

Curve D. A loss which appears to be independent of magnetic field over a wide range and can be related to the dielectric loss tangent of the material, hence called dielectric loss.

Curve E. Higher order modes causing erratic variations in loss.

Curve F. Double peaks due to "Cavity Resonances".

Qualitative and semiquantitative explanations have been developed to explain all of these phenomena. Some of them follow from a simple extension of the plane wave theory and the rest are based upon considerations of the special case of a partially filled waveguide.

Curve A, Fig. 5

Associated with the precessional resonance there is a damping term by which power is dissipated in the lattice. The exact nature of this damping term is not fully understood, and measured line widths are always greater than those predicted by present theory. Nevertheless we have at our disposal empirical damping constants which can be used to predict resonance absorption losses as was done in the calculation of the curves of Figs. 1 and 4.

These apply, however, only to small ellipsoidal samples which are ground from single crystal ferrites. In polycrystalline ferrites the absorption line is generally broader for three reasons, namely; crystalline anisotropy, strain anisotropy and varying internal demagnetizing fields due to the variety of shapes of the constituent crystallites.

Many ferrites have a high crystalline anisotropy which behaves in

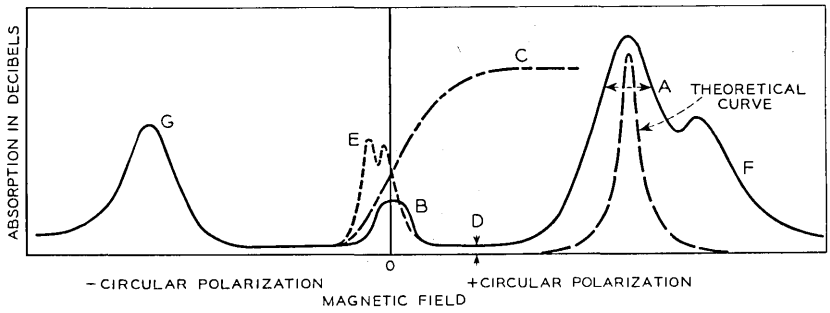


FIG. 5 — Typical loss characteristics encountered in the measurement of various ferrite samples in cylindrical waveguides with longitudinal static field.

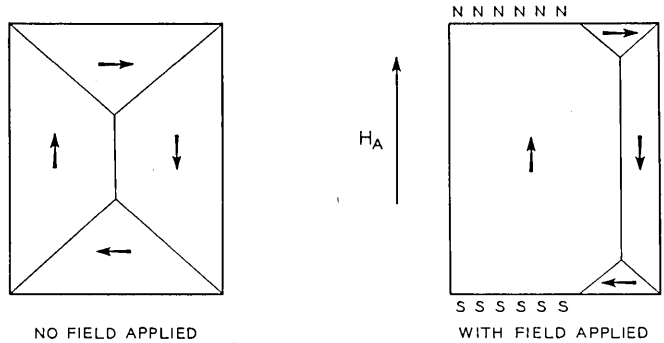


FIG. 6 — A typical domain wall pattern showing the movement of the wall in response to the application of an external magnetic field.

many ways like an internal field tending to keep the magnetization of the constituent crystallites along one of the axes of easy magnetization. In most ferrites there are four such axes and since the magnetization can be in either direction along any one of these, there are eight directions of easy magnetization. When a field is applied the effective internal field is roughly the vector sum of the applied field and the anisotropy field associated with the nearest axis. In a polycrystalline ferrite composed of many randomly oriented crystallites it is evident, therefore, that the internal field varies from point to point in the body so that the resonance absorption line is broadened by an amount proportional to the magnitude of the anisotropy field. A similar broadening can arise from magnetostriction due to fields arising from varying strains throughout the polycrystalline ferrite, and non-uniform internal demagnetizing fields due to the shape of the constituent particles or crystallites can likewise broaden the resonance line. Since a broad resonance line results in ellipticity of the transmitted wave it is desirable to use a ferrite having low anisotropy.

Curve B, Fig. 5

Frequently a loss is observed at low fields as indicated in Fig. 5 by Curve B. Neglecting waveguide effects this hump is symmetrical[3] so that it evidently depends on a phenomenon which affects both circular components equally. It is generally agreed that it depends upon the existence of domain walls within the material since it usually disappears as soon as the body is magnetized. However, there is some question as to the specific mechanism involved.

[3] Fox and Weiss, Revs. Mod. Phys., **25**, p. 262, Jan., 1953.

A ferromagnetic crystal consists entirely of regions called domains which are completely magnetized along one of the directions of easy magnetization. In general the direction of magnetization of these domains is varied in an orderly manner as shown in Fig. 6 so that the energy of the crystal as a whole is a minimum. In the region between adjacent domains there is a (usually) narrow wall in which the magnetization goes through a gradual change in direction from that of one domain to that of the other. When an external field is applied the magnetization of the crystal is increased by the growth of some domains at the expense of their neighbors. When the crystal is saturated substantially all of the walls have disappeared and the material behaves as a single large domain.

There are currently two proposed mechanisms by which these domain walls could cause a loss at low fields. Becker and Döring[4] have shown that there can be associated with the motion of a domain wall either relaxation or resonance frequencies. Galt[5] has measured relaxations in a single crystal of magnetite at 3,000 cps and in a single crystal of nickel ferrite at approximately 2.5 mc and has presented a rather convincing argument that these are due to domain wall motion. In general the *relaxation* frequency would be expected to occur far below the microwave frequencies, but resonances could conceivably occur at microwave frequencies and could be quite broad. Until recently no other theory had been advanced which would explain the losses so often observed at low fields, and these were, therefore, attributed to a high frequency domain wall resonance.

There is a more satisfactory explanation which has recently been stated in different ways by Rado[6] and by Smit and Polder.[7] Rado has observed a resonance absorption in the microwave region with zero applied field and has shown from temperature dependence that the frequency of this resonance depends upon the saturation magnetization and the crystalline anisotropy of the ferrite. Smit and Polder have presented a model by which we can see how both of these quantities can enter to produce a loss at low fields.[7] We consider an ellipsoidal crystallite as shown in Fig. 7. The domain structure shown is one which could exist in some crystallites in a polycrystalline ferrite.

The magnetization in domains numbered 1 will respond to right circular polarization and the others to left circular. In other words a wave rotating clockwise is positive circularly polarized in domains one while a

[4] Becker and Döring, *Ferromagnetismus*, Springer, Berlin, 1939.
[5] J. K. Galt, Phys. Rev., **85**, Feb. 15, 1952.
[6] G. T. Rado, R. W. Wright, et al., Phys. Rev., Nov. 1952.
[7] D. Polder and J. Smit, Revs. Mod. Phys., **25**, pp. 89–90, Jan. 1953.

wave rotating *counterclockwise* is positive circularly polarized in domains numbered 2. In the absence of any other effects the resonance frequency of all of these domains would be determined by the anisotropy field. However, if we excite both circular polarizations simultaneously and if the relative phase of the two circular polarizations is as shown in Fig. 7(a) poles will be set up at the domain walls as indicated in the figure. The demagnetizing fields associated with these will cause the resonance for both circular components to occur at a frequency given by:

$$f = \gamma H_{\text{eff}} \cong \gamma M_z \tag{13}$$

On the other hand, if the phase of the two circular polarizations is as shown in Fig. 7(b), no poles will be set up on the walls and the resonance will be determined primarily by the anisotropy field.

$$f = \gamma H_{\text{eff}} \cong \gamma H_{\text{anis}} \tag{14}$$

These two examples of the relative phase of the circular waves correspond to linear polarizations in the x and y directions respectively.

This simplified derivation gives the maximum and minimum frequencies at which resonances can occur. In a material containing a large number of randomly shaped and randomly oriented crystallites, resonances can occur at all frequencies between these limits. Most ferrites have values of M_s between 80,000 and 240,000 and anisotropy fields which probably range from 8000 to 80,000 amp. turns/meter with per-

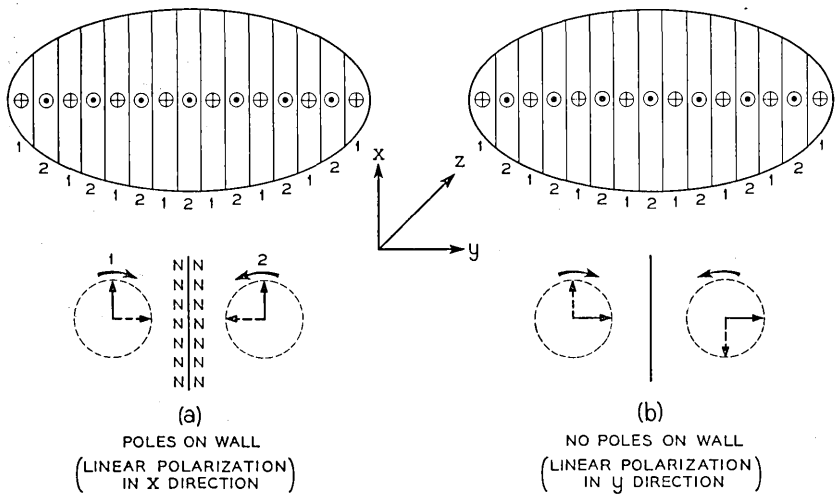


FIG. 7 — Model used by Smit and Polder to illustrate their theory for the "low-field loss."

haps a few which are higher still. Therefore, we have as typical frequencies

$$f_1 = \gamma H_{\text{anis}} \cong 300 \text{ to } 3{,}000 \text{ mc} \tag{15}$$

and

$$f_2 = \gamma M_s \cong 3{,}000 \text{ to } 9{,}000 \text{ mc}. \tag{16}$$

It is evident that at 9,000 mc only that loss associated with M_s will contribute to the "Low Field Loss" and since the mechanism depends upon the existence of domain walls it will disappear when the material is saturated in agreement with our observations.

Curve C, Fig. 5

Some ferrites exhibit a low-field loss which disappears at saturation for only the negative component and increases for the positive component. This is thought to be due to an effective anisotropy field in the material. In order for such behavior to be present at 9,000 mc, however, the internal field must be of the order of 240,000 ampere turns per meter. While such a value of crystalline anisotropy might be found in cobalt or other high anisotropy ferrites it appears to be somewhat too high for a nickel-zinc ferrite such as that in which this characteristic was first observed. However, high internal fields could result from demagnetizing effects similar to those discussed under item B but differing in that the poles are set up on nonmagnetic grain boundaries instead of domain walls. These, of course, would persist when the body is saturated, but there would be loss for only one circular component inasmuch as all of the crystallites are then magnetized in the same direction. Such a loss characteristic can be quite useful where one wishes to absorb one circular component selectively without the necessity for applying a large dc field.

Curve D, Fig. 5

Dielectric losses are present in all of the ferrites which have been made to date, although in some materials this loss is very low. Low dc conductivity in itself is not a sufficient criterion of the dielectric properties of a material as some ferrites appear to consist of conducting regions surrounded by an insulating matrix, and these have fairly high loss tangents at microwave frequencies.

The major mechanism of dielectric loss involves the exchange of electrons between ions in the crystal lattice. It has been found that the presence of ions of the same metal in different valence states on the same

lattice site gives rise to high conductivity and hence high dielectric loss. Conversely, when a ferrite is carefully prepared so that all of the constituents are present in exactly stoichiometric porportion, and when the possibility of multiple valence states is eliminated the conductivity is very low. To illustrate this point a series of measurements is reported in which the iron content of the ferrites was carefully varied about stoichiometry in a nickel-zinc ferrite. These measurments are discussed at a later point in this paper.

A ferrite is made by reacting a mixture of metallic oxides at a temperatures below the melting point of these oxides. As the oxides react a new crystal structure is evolved in which the metallic ions occupy positions in the interstices of a close-packed oxygen lattice. There is a very well authenticated theory due to Neel[8] explaining the way in which the spin orientations of the ions are distributed in the two types of lattice site which exist in the Spinel oxygen lattice. Whenever metal ions in more than one valence state occupy the same type of site, e.g., the octahedral position, there is a possibility for the easy transfer of an electron from one to the other since the crystal structure is unchanged by the transfer.[8] In the case of nickel ferrite which has the composition $NiOFe_2O_3$ an excess of iron will tend to replace some nickel atoms by entering the lattice in the divalent state. Since the remainder of the iron is trivalent, comparatively high conductivity is observed. The problem of producing ferrites with extremely low dielectric losses appears to be fairly well understood and is progressing satisfactorily. By choosing the proper set of metal ions to insure the absence of multiple valence states and by maintaining the proper oxygen stoichiometry one may be able to achieve loss tangents as low as 0.001. The subject of dielectric losses is well covered in the literature.[9, 10]

Curves E, F and G, Fig. 5

The loss mechanisms indicated in Fig. 5 by Curves E and F all arise from the particular behavior of ferrites in waveguides as differentiated from the plane wave theory.

For example, the erratic behavior indicated by Curve E has been shown to be due to the presence of higher order modes in the ferrite region in a waveguide, and the subsidiary hump on the absorption Curve F has been shown to be a "cavity resonance" which is strongly dependent

[8] L. Neel, Physica, **16**, pp. 350–53, 1950, and Zeit. Anorg. Chem., **262**, pp. 175–184, 1950.

[9] E. J. W. Verwey and J. H. DeBoer, Rec. des Travaus Chemiques des Pays-Bas, **55**, pp. 531–54, 1936.

[10] E. J. W. Verwey et al., Phillips Res. Reps., **5**, pp. 173–187, 1950.

upon the diameter of the ferrite cylinder and upon the guide wavelengths but not upon the length of the cylinder.

Other waveguide effects causing anomalous loss behavior, such as shown by curve G, have been discussed by Fox and Weiss[11] and will be treated by them in greater detail in a forthcoming publication.

In order to discuss these effects more fully we must examine the modifications of the plane wave theory which must be made to explain the behavior of a ferrite in a waveguide.

WAVEGUIDE THEORY, LONGITUDINAL FIELD

When a piece of ferrite is placed in a waveguide and magnetized it is necessary to modify the foregoing plane wave theory to describe the behavior of a wave passing through the ferrite. Because of the anisotropic nature of the magnetized ferrite it is necessary to obtain a solution to the specific problem of the waveguide containing the ferrite. When the magnetization of the ferrite precesses about the applied dc field it sets up components of **h** which do not exist in any of the classical modes, and unless one can deal with small perturbations the solution becomes quite involved.

The modes which can exist in the ferrite will often resemble the classical modes so that for convenience we will refer to them as modified TE or TM modes. Suhl and Walker[12] have obtained solutions for the case of a cylindrical waveguide completely filled with a magnetized ferrite, and they have shown that the modified dominant TE_{11} mode behaves much like the plane wave in the region of small fields but that the behavior of the TM modes cannot be approximated by a simple extension of the plane wave theory. A waveguide large enough to support the dominant mode when filled with air will, when filled with ferrite, support three or four higher order modes including some of the modified TM modes. In this case, it is possible to have several present at the same time with the result that observations of rotation, loss and ellipticity are almost impossible to interpret. Accordingly, we should reduce the size of the waveguide in the ferrite-filled region, and this involves the creation of discontinuities in the waveguide. This is not always necessary, however, because higher order modes will not always be set up in the ferrite even though the waveguide is large enough to propagate them. If care is taken to avoid geometries which favor a given mode the prob-

[11] A. G. Fox and M. T. Weiss, Revs. Mod. Phys., **25**, p. 262, Jan., 1953.

[12] A preliminary report of this work has been published in the form of a Letter to the Editor by H. Suhl and L. R. Walker in Phys. Rev., **86**, p. 122, 1952. A more detailed account is scheduled for publication in the J. Appl. Phys.

ability of its occurrence will be greatly reduced. In particular it has been found that a flat-ended cylinder completely filling the waveguide can be introduced into the full-sized waveguide without mode complications, but Fox and Weiss[13] have shown that putting conical tapers on the ends will favor the establishment of the modified TM_{11} mode. In most applications of the Faraday effect the ferrite element is in the form of a very thin pencil at the center of the waveguide so that the mode problem is greatly simplified, but in order to obtain quantitative fundamental information about ferrites themselves it is often necessary to work with a completely filled waveguide. In such cases considerable care must be taken to insure the validity of the measurements.

One method of making impedance measurements which has been used successfully by H. Suhl is to cut a shallow longitudinal slot in the cylinder of ferrite and to make standing wave measurements directly in the medium. Because the slotted section is filled with ferrite the size of the waveguide can be reduced to insure the presence of a single mode. This is restricted to unmagnetized materials as rotation of the plane of polarization would result in radiation by the slot.

Another suggested procedure is to make a transformer from full-size rectangular waveguide to circular waveguide of diameter equal to $d/\sqrt{\epsilon}$ where d is the diameter of a dominant-mode air filled pipe and ϵ is the relative dielectric constant of the ferrite. This transformer can be treated as a four terminal impedance transformer and its network impedances can be determined by measurement. Impedances measured in the air filled guide will have to be transformed through this network to obtain the true impedances of the ferrite-filled guide, but this can be done if the need for the measurements warrants such effort.

An exact solution of the partially filled waveguide is considerably more difficult to obtain than the solution for a completely filled waveguide. Yet this geometry is the one usually used in most practical applications of the Faraday effect. In the absence of an exact solution one must develop simple physical explanations based upon plane wave theory plus intuition for numerous observed phenomena. A theory for the partially filled longitudinally magnetized waveguide can easily be developed from two simple observations. First, we consider the circular components of the wave separately and observe that each sees an effective scalar permeability which is a weighted average of the permeability of the pencil for that component and that of the surrounding medium, and second we postulate that a small enough pencil will not act as a dielectric rod waveguide and will merely create a small perturbation of

[13] A. G. Fox and M. T. Weiss, Revs. Mod. Phys., **25**, p. 262, Jan., 1953.

the original mode. When the above assumptions are valid the plane wave theory can be extended easily to explain loaded waveguide behavior. In the plane wave case it was shown that a negative effective permeability results in attenuation of the positive circularly polarized wave. Quite a different result is observed in waveguides containing very small cylinders of ferrite. It appears that if the cylinder is small enough not to act as a dielectric waveguide then the negative permeability inside the rod simply is averaged with the permeability of the surrounding region so that the rotation curve (which depends upon the difference between the square roots of the permeabilities seen by the two circular components of the wave) follows the dispersion curve of the permeability of the positive circularly polarized wave, even following the pattern of the permeability when it is negative.

In Fig. 8 are shown measured curves of the rotation of the wave and the absorption of the positive circularly polarized component of the wave as functions of applied dc field for comparison with Figs. 1 and 2. To amplify our arguments we point out that the propagation constant in a waveguide containing a very small pencil of ferrite is of the form:

$$\beta_{\pm}^2 = \beta_e^2 + \mathrm{A}\frac{r_1^2}{r_0^2}\left[\frac{\mu_{\pm} - 1}{\mu_{\pm} + 1}\beta_e^2 - \beta_0^2\frac{\epsilon - 1}{\epsilon + 1}\right] \tag{17}$$

where β_e is the propagation constant of the empty guide
r_1 is the radius of the ferrite cylinder
r_0 is the radius of the waveguide
ϵ is the dielectric constant of the ferrite
and $\mu_{\pm}$ are the effective complex permeabilities ($\mu \pm k$)
A is a constant $\cong 3.2$

We see that the expression within the brackets is finite for all values of $\mu_{\pm}$ except $\mu_+ = -1 + j0$. Accordingly if the damping parameter (or line width) is large enough to insure that μ_+ can never take on this value, there will always be a cylinder radius, r', for which the perturbation term is small relative to β_e^2. However, the cylinder diameter does not have to be very large before the above arguments fail and the rotation and loss behavior become quite different.

The cutoff wavelength of the TE_{11} mode in round guide is

$$\lambda_c = 0.1708\, d\, \sqrt{\mu\epsilon}$$

where d is the diameter in centimeters of the waveguide and μ and ϵ are the effective relative permeability and dielectric constant of the medium contained therein. When a cylinder of ferrite is placed at the center of

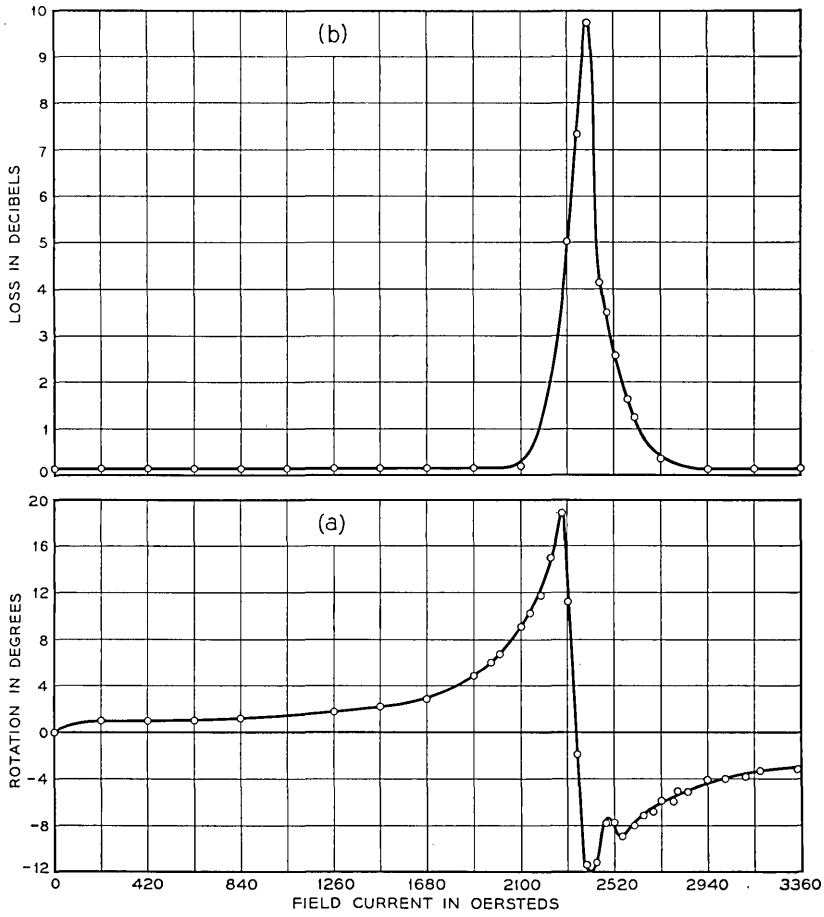


FIG. 8 — Rotation of the plane of polarization and absorption of the positive component as functions of applied dc field for a wave propagating through a waveguide containing a very small cylinder of ferrite. The inflection of 2,400 oersteds is due to a "cavity resonance" as discussed later.

the waveguide the cutoff wavelength is increased because the effective value of ϵ is increased. If the ferrite cylinder is quite small this alteration will be unimportant, but if the diameter of the rod exceeds about one-quarter that of the air filled dominant mode guide this mechanism can lead to the existence of higher order modes for the negative component. When this happens the plane wave theory obviously cannot be expected to apply, for the presence of multiple modes in the propagation of either component will introduce an additional variable not present in the plane

wave theory. Experimentally one will observe very erratic and frequency-dependent behavior under these conditions.

WAVEGUIDE THEORY — TRANSVERSE FIELD

A waveguide, either round or rectangular, filled with ferrite and magnetized by a field parallel to the *electric* vector of the dominant mode will exhibit a behavior qualitatively the same as described in the plane wave theory of the transverse field. In fact, it has been shown that in a rectangular guide all of the TE_{on} modes can exist with only slight modification.[14] That this result is probable may be seen from the fact that the precessing magnetization vector sets up components of **h** in the x and y directions when the applied field is in the z direction, and both of these components normally exist in the TE_{on} modes. The primary modification of the mode arises from RF demagnetizing fields in the ferrite. Because of this modification it is extremely difficult to match the boundary conditions for normal incidence at an interface between the ferrite and air in the waveguide. An infinite series of modes is actually required, but in practice the mismatch due to magnetic effects is usually not very large. If one matches the dielectric constant by means of tapered dielectric horns the remaining mismatch is slight except where μ_{eff} approaches zero and at resonance.

While the completely filled waveguide magnetized by a transverse magnetic field parallel to the electric vector of the wave will exhibit a reciprocal behavior in respect to phase change and attenuation, an interesting and potentially useful modification of these effects occurs when a small piece of ferrite is located asymmetrically in a waveguide. Chait and Sakiotis[15] of the Naval Research Laboratory and Turner of the Holmdel laboratory of Bell Telephone Laboratories have independently observed a phase shift which is dependent upon the direction of propagation of the wave, and a simple explanation of this effect has been made by Turner[16] and by Kales.[15] Suhl and Kales[15] have shown the theoretical validity of this explanation. The idea can be demonstrated by consideration of the field configuration shown in Fig. 9.

An observer at the point P will see an **h** field which is elliptically polarized in a plane normal to the direction of H_A. The sense of the rotation of the larger circular component of the ellipse will depend upon the direction of propagation of the wave. Thus for *one* direction the major

[14] A. A. van Trier, Paper presented orally at meeting of Amer. Phys. Soc., Washington, D. C., April, 1952.

[15] M. L. Kales, H. N. Chait and N. G. Sakiotis, Letter to the Editor, J. Appl. Phys., June, 1953.

[16] E. H. Turner, Letter to the Editor, Proc. I. R. E., **41**, p. 937, 1953.

part will be a *positive* circular polarization with respect to the applied H_A and will experience a decrease in μ' and for the other direction of propagation the **h** field will be primarily *negative* circularly polarized and the wave will experience an increase in μ'. Since the **h** field is linearly polarized in the transverse plane at the center of the guide and is linear in the longitudinal direction at the edge of the guide it is evident that there is a point in between where the differential phase shift is maximum. Suhl has shown that this point always occurs halfway between the guide wall and the center regardless of the proximity of cut-off.

Obviously one has merely to adjust the length of the sample and the strength of the magnetic field so that the differential phase shift is 180° and he will have a gyrator. Then it is an easy matter to design a circulator, isolator, or any of the numerous devices depending upon the gyrator action.[17]

EXPERIMENTAL PROCEDURES AND RESULTS

In order to verify the theory and to determine the optimum performance obtainable in microwave devices employing the Faraday Rotation an extensive measurement program has been set up. Information of both a fundamental and of a practical nature is obtained through a variety of measurements. From a practical point of view we are interested

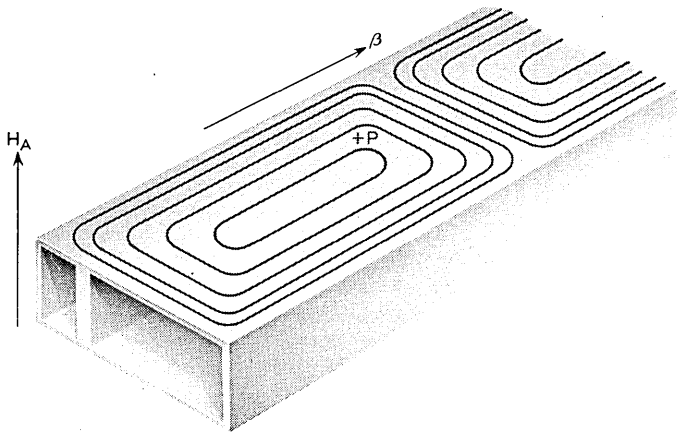


FIG. 9 — Rectangular waveguide containing assymetrically located ferrite slab. Magnetic lines of force are shown on the top wall of the guide.

[17] C. L. Hogan, The Ferromagnetic Faraday Effect at Microwave Frequencies and Its Applications — The Microwave Gyrator, B.S.T.J., **31**, pp. 1–31, Jan., 1952.

in obtaining a ferrite which gives the maximum rotation per unit loss, and at the same time we are interested in obtaining such fundamental information as the mechanisms of loss, the effect of composition, and the relationship between the various physical properties of the ferrite crystals and their microwave behavior. As a corollary interest we wish to relate the microwave behavior of ferrites to their low frequency performance. For convenience most of the microwave measurements have been made in the 9,000 mc X-Band region, but 4,000, 24,000 and 48,000 mc measurements have been carried out by others in Bell Laboratories and some of their findings will be reported here for comparison with those obtained at X-Band.

In the most common procedure the Faraday Rotation and attenuation of a linearly polarized wave and the ellipticity of the resultant wave are measured as a function of the intensity of the applied longitudinal magnetic field. The experimental equipment used in these measurements is shown in Fig. 10(a). A variety of sample shapes ranging from cylinders completely filling the waveguide to very small cylinders suspended along the axis have been measured. From these data the absorption of the positive circular component may be obtained if one assumes that the attenuation of the negative component remains constant as predicted in the plane wave theory. This is not always the case, however, due to the mode configurations sometimes present, and for this reason an alternate

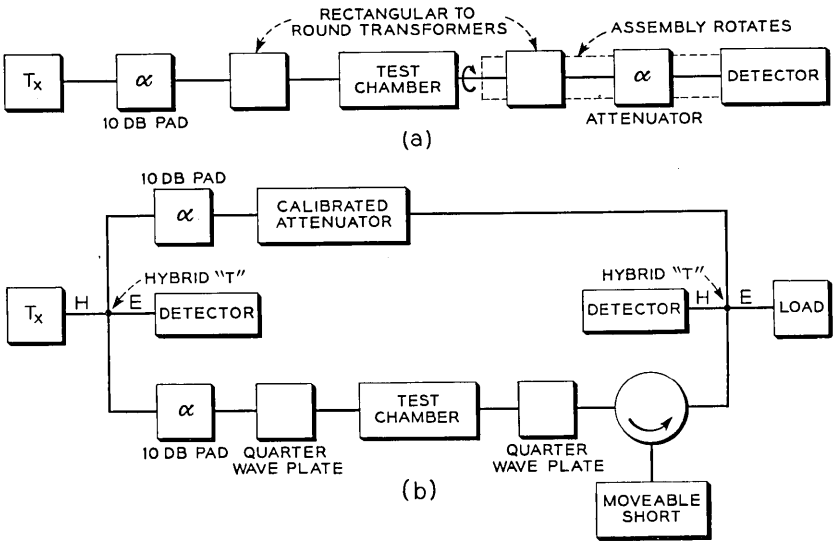


FIG. 10 — Block diagrams of the two most commonly used measuring set-ups.

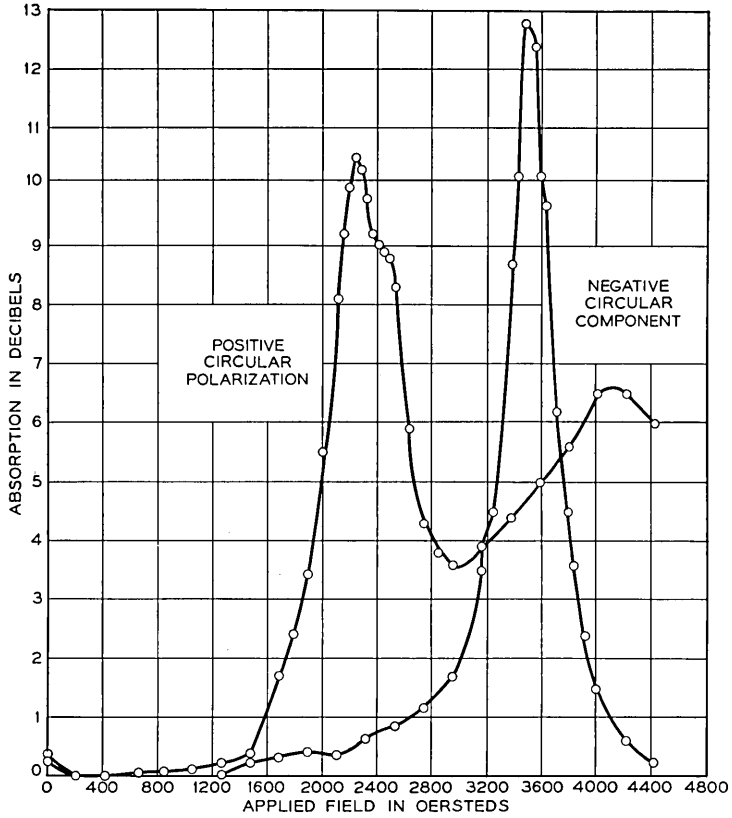


FIG. 11 — Absorption versus applied field for positive and negative circular polarizations. Ferrite cylinder diameter was 0.125".

procedure is often used. This consists simply of exciting the circularly polarized components separately and measuring the phase shift and attenuation of each of these components with the equipment shown in Fig. 10(b). There is no ambiguity in these results, and the equivalent rotation can easily be calculated from the phase data.

Because Polder's permeability matrix is made linear by approximations it is desirable to verify that the measurements made with circular polarization can be superposed to produce the same result as would be obtained using linear polarization.[18] For this purpose we choose a ferrite sample which exhibits a fairly complicated loss curve for both circular components. The absorption characteristics for both circular polariza-

[18] An experiment of this type was first performed by J. P. Schafer of Bell Telephone Laboratories in 1951 to establish the veracity of observed Faraday Rotations.

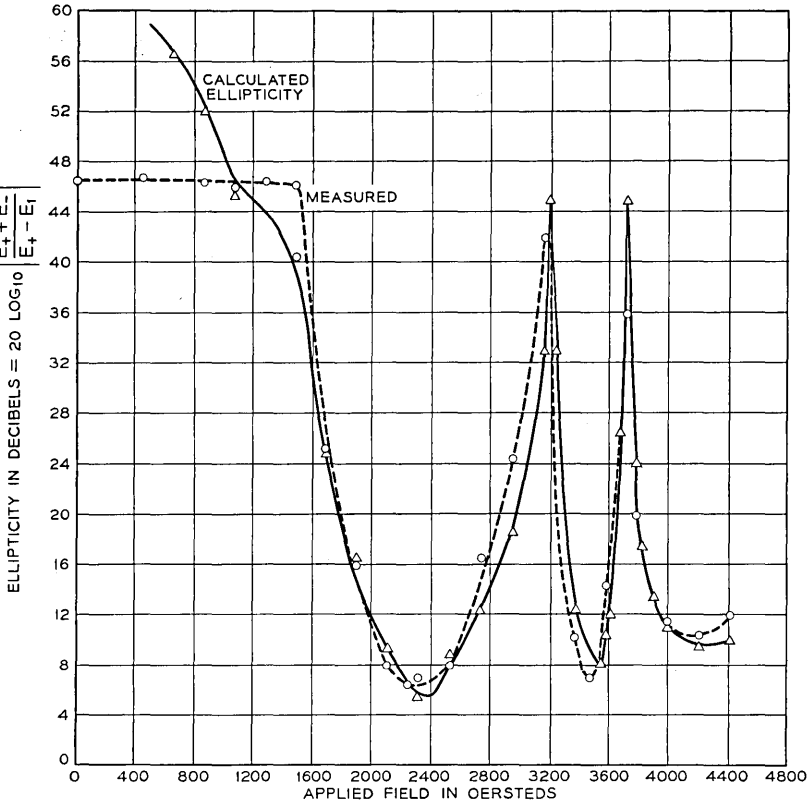


FIG. 12 — Ellipticity versus applied field as observed and as calculated from the data of Fig. 11.

tions are shown in Fig. 11. The resultant ellipticity in db is calculated from the relation

$$\Delta = 20 \log_{10} \left| \frac{E_+ + E_-}{E_+ - E_-} \right| \tag{18}$$

where $E_\pm$ are the rms values of the circular components. This result is shown by the solid line in Fig. 12, and the ellipticity as measured by the set-up of Fig. 10(a) is shown by the dashed line. The agreement between these two curves is quite good if we consider that the accuracy of measuring each circular component is no better than ± 0.05 db so that in the regions where the two circular components are actually equal the calculated ellipticity can vary from infinity down to 52 db, while the signal to noise ratio of the equipment limits the measured ellipticity to 50 db.

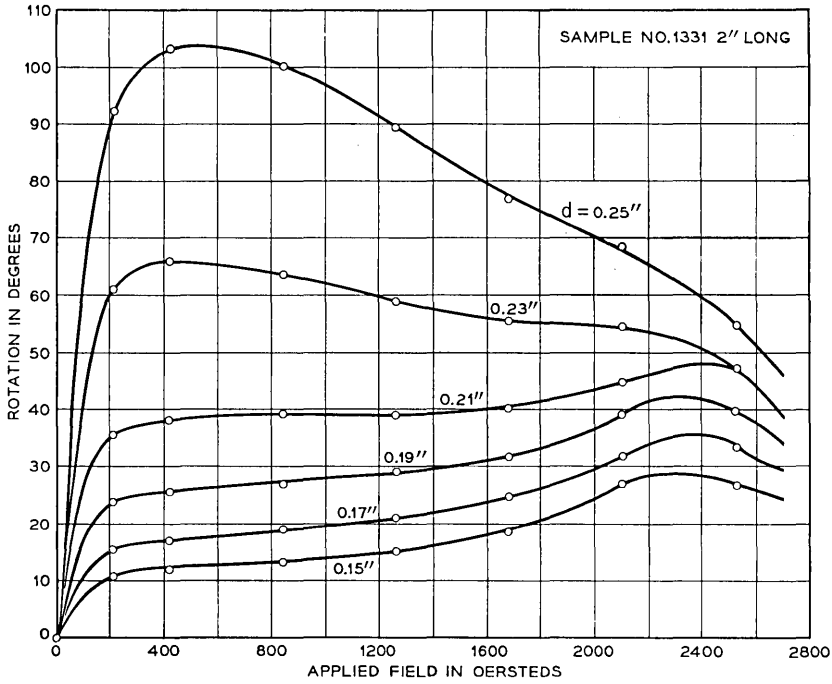


FIG. 13 — Dependence of rotation upon ferrite cylinder diameter.

To illustrate the effect of the diameter of the ferrite specimen upon the Faraday Rotation the following experiment was performed. A one-quarter inch cylinder suitably tapered on both ends was measured in the rotation measuring equipment and then ground down in steps of about 0.020" and remeasured after each cut. The particular ferrite chosen was one having very low dielectric loss and a dependence upon sample diameter which is typical of many ferrites. The data are shown in Fig. 13. At the largest diameter we observe a rotation which agrees with neither the plane wave theory nor the small perturbation approach. This behavior is evidently due to the presence of a higher order mode of propagation in the section of waveguide containing the ferrite cylinder.

Higher order mode effects are undesirable from the standpoint of bandwidth as the rotation produced thereby is much more frequency sensitive than that obtained from the TE_{11} mode. This is evident from consideration of the fact that the partially filled waveguide is much closer to the cutoff for the higher order modes than for the dominant mode, and we will show that frequency dependence is increased as waveguide approaches cut-off for a given mode.

BANDWIDTH OF THE FARADAY ROTATION

Some consideration has been given to means of increasing the bandwidth over which the Faraday Rotation is relatively constant. Two possible ways of broadbanding the effect can be suggested on the basis of measurements and theory.

In the infinite medium-plane wave theory the Faraday Rotation is shown to be totally independent of frequency everywhere far above the ferromagnetic resonance frequency. One way of explaining this lack of dependence upon frequency is to observe that the rotation per unit wavelength *decreases* with increasing frequency while the number of wavelengths per unit length *increases* at the same rate so that the rotation per unit length remains constant. In a waveguide there are two effects which cause the rotation to be frequency dependent. In the first place the guide wavelength is not linearly related to frequency and secondly, where thin pencils of ferrite are used, the radial distribution of field varies with frequency in such a way as to add to the frequency dependence.

By surrounding the ferrite element with a material of the same dielectric constant as that of the ferrite the guide wavelength will be reduced to approximately one-fifth the cut-off wavelength and will be almost linearly related to the frequency in this region. Furthermore, the radial distribution will not change appreciably with frequency because, neglecting the difference in permeabilities, the waveguide is now filled with a uniform dielectric. On this same basis there should be no tendency for the structure to set up higher order modes, so that if care is taken in the design of the transitions from circular to rectangular waveguide, the higher order modes can be avoided.

While the above approach would give very good bandwidth a lack of very low-loss dielectrics having the proper values of dielectric constant limits its usefulness. One of the best dielectrics in this range is Micalex K, and it has a loss tangent of approximately 0.001 which would produce a total loss of several tenths of a db in a length of two or three inches.

The bandwidth can be increased in another way which requires no special dielectric materials. Suhl has shown that when the ferrite element is not perfectly matched to the waveguide the resulting multiple reflections can enhance or detract from the inherent rotation. When the element is an integral number of half wavelengths long the resulting rotation is maximum and when the element is an odd number of quarter-wavelengths long the rotation is minimum. Since the rotation of a perfectly matched cylinder increases with increasing frequency we must choose a length for the unmatched element such that it is an integral

number of half-wavelengths long at the low end of the frequency band in order to increase the rotation at that point. The ferrite chamber may be designed as a resonant cavity very tightly coupled so that over the band the input impedance satisfies the matching requirements even though there are multiple reflections inside the cavity.

Some ellipticity of the resultant wave will be introduced by this method as the wave will be linearly polarized only at those frequencies at which the ferrite element is an integral number of quarter-wavelengths long. However, if only a small correction is applied in this way the ellipticity will be tolerable.

The first and second broadbanding techniques may be combined in a manner which was discovered in the course of our measurements to determine bandwidth. Consider a circular waveguide such that at the low end of the band the wave is just slightly above cut-off. Then the wave impedance in the air-filled pipe will be high. If now the pencil of ferrite is supported in a polystyrene holder, the wave impedance in the ferrite region is much lower than that in the air filled region and multiple reflections will be set up. In order to reduce the magnitude of the reflections to achieve just the right degree of correction each end of the polystyrene can be tapered at the proper angle to give optimum correction. The polystyrene alone gives some improvement in bandwidth according to the arguments first submitted, and the internal reflections provide the rest of the compensation. The data shown in Fig. 14 show the variation of rotation with frequency obtained in this manner. The dotted extension of the low end of the curve indicates the expected rotation in the absence of the correction. By means of this compensation we are able to restrict the variation in rotation to ± 4 per cent over a band wider than 15 per cent. Over this band the ellipticity defined as the ratio in db of the maximum to the minimum field strength was over 50 db which corresponds to perfectly linear polarization within our ability to measure it. As a further advantage of this system the rotation obtained at all frequencies from the ferrite pencil is increased through the use of the polystyrene. This amplification of rotation follows directly from an extension of the plane theory.[19]

The rotation per unit length is given by the relation

$$\frac{\theta}{l} = \frac{\omega}{c}\sqrt{\epsilon}\,(\sqrt{\mu_-} - \sqrt{\mu_+}) \tag{19}$$

where $\mu_\pm$ are the effective permeabilities seen by the circularly polarized

[19] Such an amplification of rotation was first observed and studied by J. P. Schafer of Bell Telephone Laboratories at Deal, N. J.

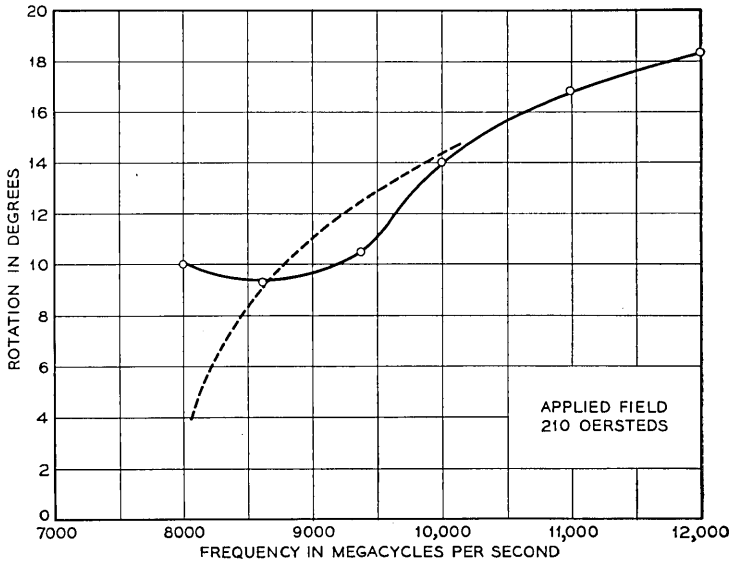


FIG. 14 — Variation of rotation with frequency showing broadbanding obtained through compensation technique.

components of the wave. The corresponding relation for the partially filled waveguide will be a transcendental expression involving $\mu_{\pm}$ and ϵ in a similar way so that arguments regarding the waveguide problem may be based upon equation 15 if we consider $\mu_{\pm}$ and ϵ to be effective values averaged over the guided mode. Thus by increasing the dielectric constant of the region surrounding the ferrite we increase the rotation by increasing the average value of ϵ. Since we also change the radial distribution in such a way as to reduce the fraction of the power contained in the ferrite the amplification in rotation will be less than would be obtained if the waveguide diameter were reduced at the same time. In Fig. 16 we show the effect of increasing the dielectric constant of the region surrounding the ferrite and the effect of a subsequent reduction in guide size. Here again we see that if very high dielectric constants were available in low loss materials a really significant improvement in performance could be obtained. Nevertheless, even the effect of the polystyrene is useful and here we suffer a loss of less than 0.1 db at X-Band.

In the discussion of the loss mechanisms the second hump on absorption Curve F in Fig. 5 was described as a "cavity resonance". While the exact mode of resonance cannot be determined except from a complete solution of the partially filled waveguide problem, we are able to show that the subsidiary hump is strongly dependent upon the diameter of

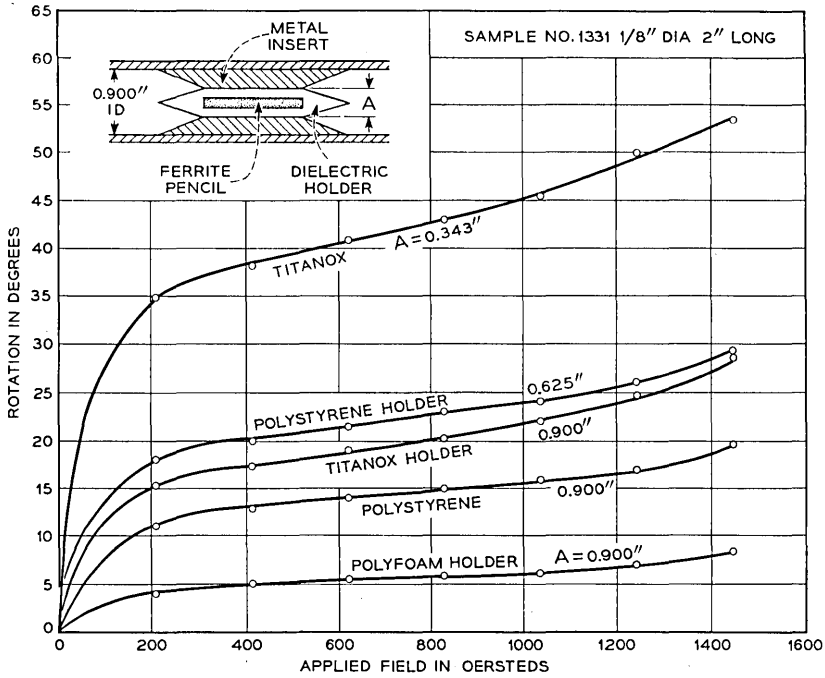


FIG. 15 — The dependence of the rotation characteristic upon the geometry and dielectric constant of the surrounding medium.

the ferrite element and upon the guide wavelength but that it is entirely independent of the total length of the element. In Fig. 16 we show the absorption of a positive circularly polarized wave as a function of applied field for several cylinder diameters. We see that the first hump remains relatively fixed in position while the second hump moves rapidly in the direction of higher fields and becomes larger in height as the diameter is increased. The first resonance is the ferromagnetic resonance, and its movement with increasing diameter is due simply to the fact that the demagnetizing factors of the cylinder are changed when the diameter is changed. The movement of the second hump can be explained if we assume that the resonance frequency of the shape resonance is given by an equation of the form

$$f = \frac{A}{d\sqrt{\mu_+ \epsilon}} \tag{20}$$

This is an equation typical of those encountered in dimensional resonances. Here A is some constant associated with the geometry of the sample and d is the cylinder diameter. At a fixed frequency, we see that

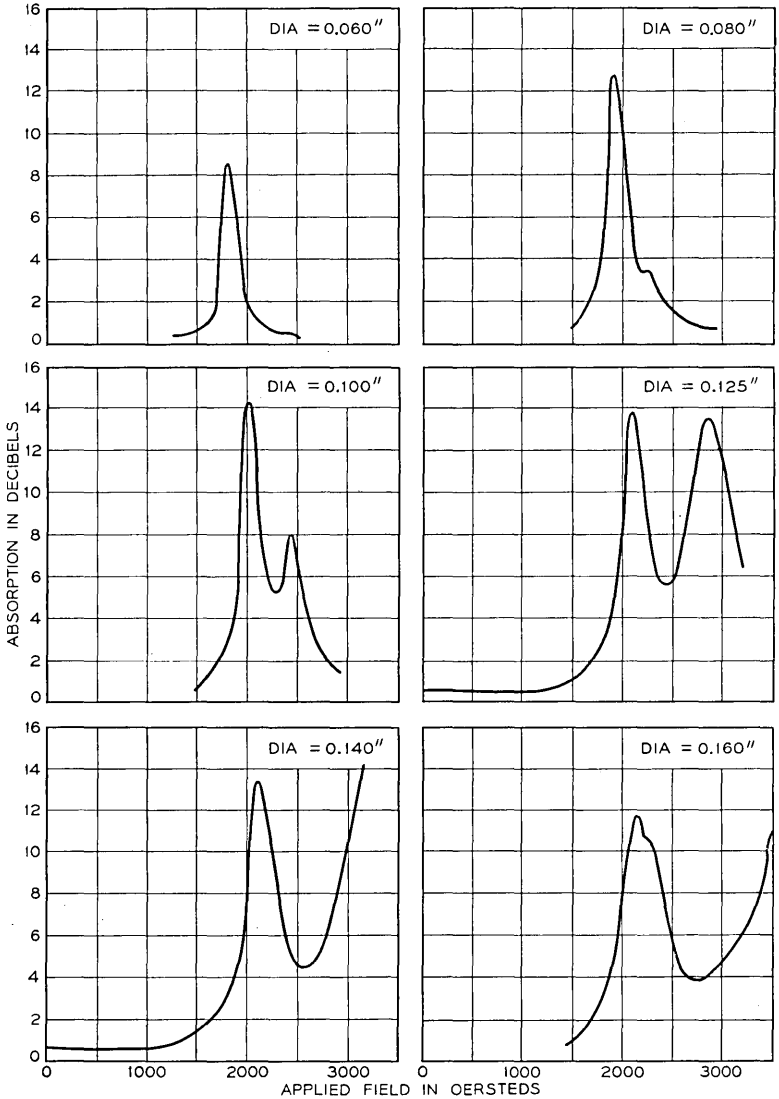


FIG. 16 — Absorption of a positive circularly polarized wave for several sample diameters showing "cavity resonances".

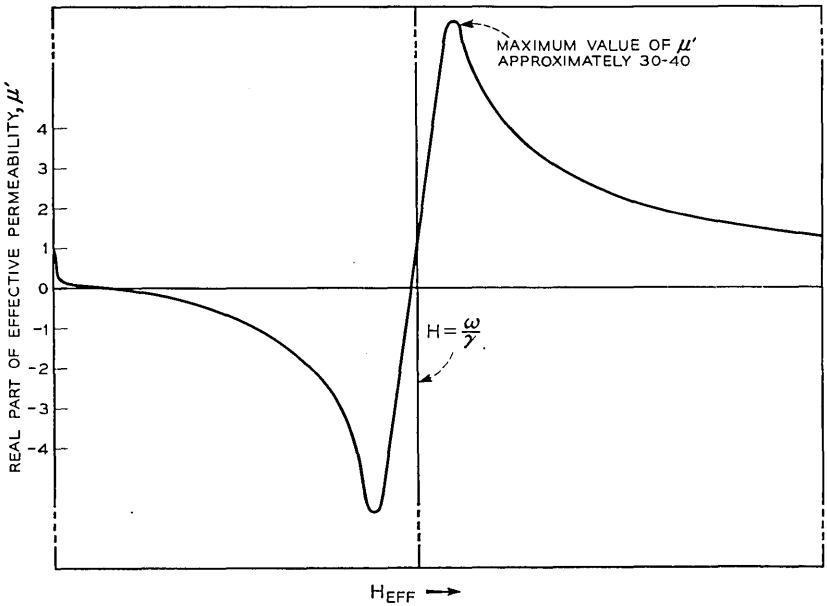


FIG. 17 — Exaggerated scale plot of the real part of the effective permeability versus effective dc magnetic field.

an increase in d must be accompanied by a decrease in μ_+' to maintain resonance. For convenience we reproduce in Fig. 17 the behavior of the real part of the effective permeability as a function of field. We see that above resonance the permeability starts at a very high value and then decreases as we move toward higher fields. Thus as we increase the diameter, d, we would expect the "cavity" resonance to move to the right, and since in doing so we go further from the ferromagnetic resonance absorption we expect the height of the peak to increase due to the fact that the loss tangent of the permeability is smaller, and hence the effective Q of the "cavity resonance" is larger. Finally we observe that the broadening of the peak which occurs in the larger diameters is due simply to the fact that the slope of the μ_+' curve decreases as we move to the right so that a larger change in dc field is required to take μ_+' through a given variation.

A similar movement of the subsidiary hump can be effected through a change in the dielectric constant of the medium surrounding the ferrite element. However, altering the total length of the ferrite in steps of 0.030″ until the total change was greater than 0.250″ produced no noticeable effect upon either the relative heights or positions of the two peaks.

FERRITE COMPOSITION AND IRON STOICHIOMETRY

There is a continuous program of measurement in which a large number of ferrites from various sources are tested at X-Band and other frequencies. In the early days of the program the most remarkable feature of the results was the tremendous variation in properties even among materials of supposedly the same composition. More recently some general trends have been observed, and it is now possible to correlate the microwave performance of certain types of ferrites with their chemical and structural nature. The nickel-zinc ferrites in particular have been studied extensively, and those having the approximate formula $Ni_{0.3}Zn_{0.7}Fe_2O_4$ are quite good for microwave applications so long as they have slightly less iron than the above formula requires. A series of measurements illustrating the effect of slight variations in iron content yielded the data shown in Figs. 18 and 19. These ferrites were prepared by F. J. Schnettler of Bell Telephone Laboratories in such a way that their bulk properties correspond very closely to their crystallite properties. The correspondence between the bulk dc conductivity and the RF

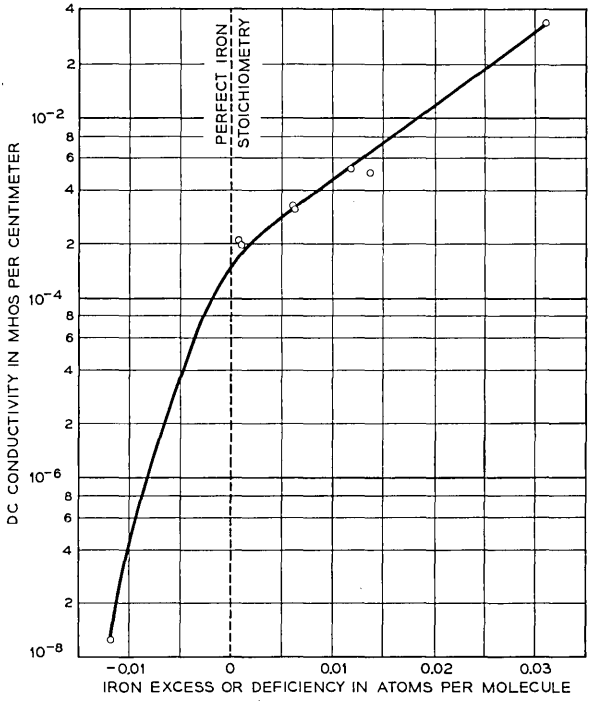


FIG. 18 — DC conductivity versus iron content of NiZn ferrite.

dielectric loss is particularly striking, indicating that the primary mechhanism of loss in these ferrites is simply ohmic conductivity loss. This series of measurements clearly shows that the presence of even a very small amount of divalent iron is undesirable in ferrites to be used at microwave frequencies.

TRANSVERSE FIELD MEASUREMENTS

Two classes of measurements are commonly made in which a transverse dc magnetic field is applied parallel to the electric vector of the TE_{01} mode in rectangular waveguide. Those measurements involving a completely filled or symetrically loaded waveguide yield primarily reciprocal effects which are useful in applications as previously mentioned. More useful, still, are the non-reciprocal effects observed when a piece of ferrite is placed asymetrically in a waveguide and transversely mag-

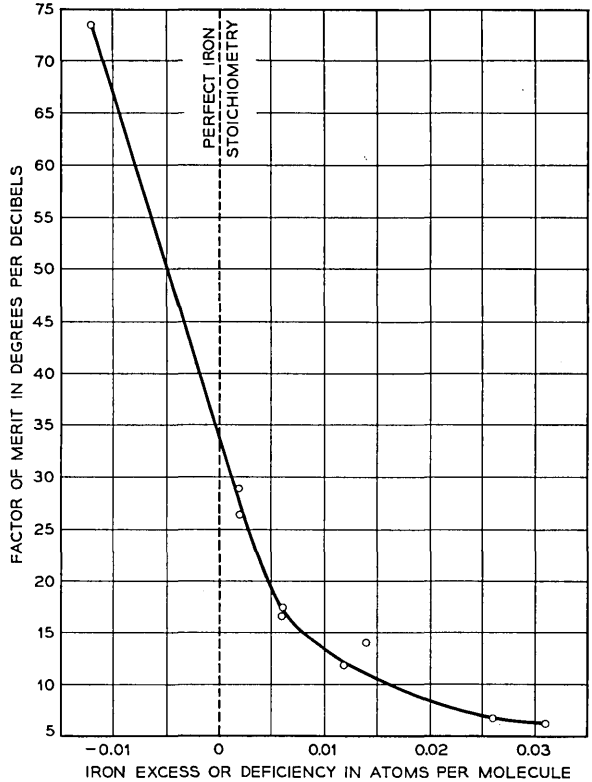


FIG. 19 — Factor of merit versus iron content of NiZn ferrite.

netized. Since these latter effects are quite new two measurements of this type are shown to illustrate the sort of performance obtainable. In the first experiment a slab of ferrite 20 mils thick was placed successively in several positions in a rectangular waveguide, and the phase shift as a function of field was measured for both directions of propagation. When the plate of ferrite is centrally located the phase shift is the same in both directions but when the ferrite is placed half way between the center and the edge of the guide a large difference in phase shift is observed. Finally when the slab of ferrite is located at the edge of the guide there is only a small difference between the positive and negative phase characteristics. This small difference will vanish as the thickness of the ferrite plate goes to zero. In Fig. 21 (a) we show the phase shift versus applied field for three positions of the ferrite slab and in Fig. 21 (b), the differential phase shift at a constant value of applied field is plotted against the position of the ferrite plate. The solid line curve taken at 9,500 mc indicates that maximum differential phase shift is obtained when the ferrite is located approximately 0.100″ from the guide wall. Suhl's prediction is that the position at which maximum differential phase shift is observed should be independent of frequency. The dotted curve in Fig. 21 (b) taken at 8,200 mc verifies this part of the prediction in that the maximum again occurs where the ferrite plate is 0.100″ from the guide wall even though the point at which **h** is circularly polarized has been shifted significantly by the change in frequency.

The second measurement was designed to measure the non-reciprocal absorption which is obtained when the strength of the dc magnetic field is adjusted so that the ferrite is at ferromagnetic resonance. In order to obtain the minimum forward loss and maximum reverse loss it is essential that the ferrite be located precisely at the point where the transverse and longitudinal components of the **h** field of the wave are equal, i.e., where the **h** field is circularly polarized in a plane perpendicular to the magnetic field. Since this condition exists at only one point in the half-waveguide the ferrite slab must be made very thin. Measurements were made in the 6000–7000 mc band in RG50 waveguide. A thin plate of "Ferramic G" was cut so as to extend from one broad wall to the other Its length was approximately 1-½″ and its thickness was originally 0.050″ and was subsequently reduced to 0.025″ and finally to 0.009″. In the last case the ferrite was so fragile that it was necessary to support it by cementing it to a ⅛-inch plate of polystyrene. For convenience the ferrite plate was fastened securely in place at a point calculated to be the point where the **h** vector is circularly polarized at the center of the band, and frequency was varied about this center frequency. At each

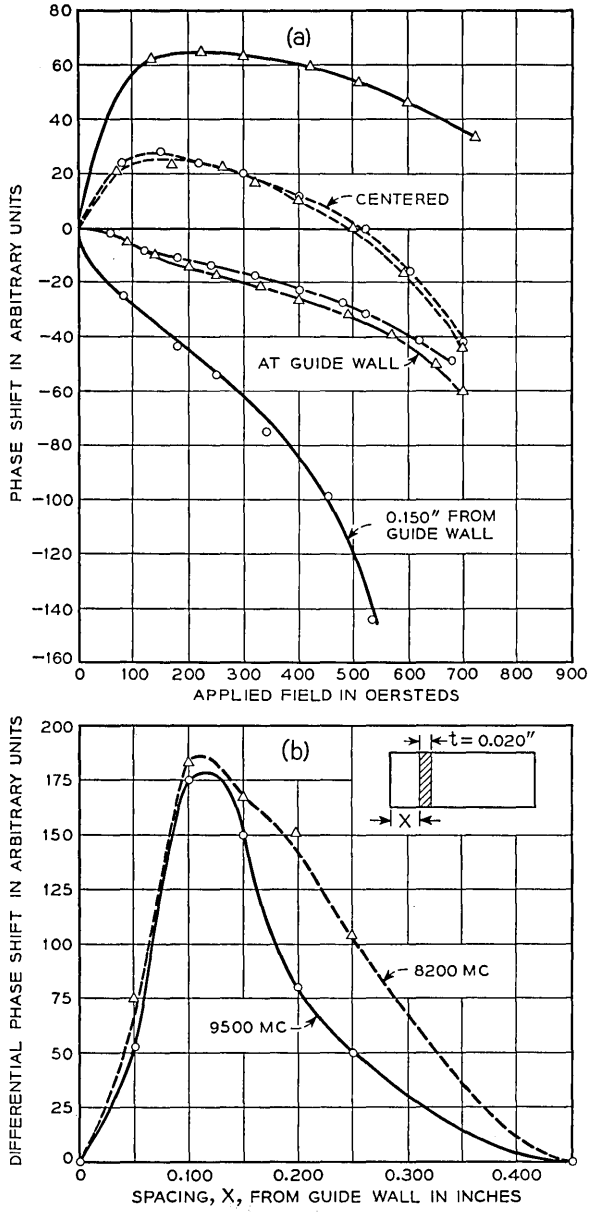


FIG. 20 — Phase shift versus applied field and differential phase shift versus position of the ferrite plate.

frequency the dc magnetic field was adjusted so that absorption was a maximum for the wave for which the absorption is the greater. The field was then reversed and the absorption was again measured. Since reversing the field is entirely equivalent to reversing the direction of propagation we obtain in this way the insertion loss for both directions of propagation. Data taken in this manner are shown in Fig. 22 along with a sketch of the geometry employed. The VSWR was measured in the case of the thinnest plate and was found to be less than 1.05 to 1 over the band. The variations in the dotted curve, however, are probably due to reflections from the ends of the sample.

One must bear in mind that as the frequency is changed, the field required for ferromagnetic resonance is changed so that these data do not give a true index of the bandwidth of the device. This must be measured at a constant value of field. However, if we had used a ferrite

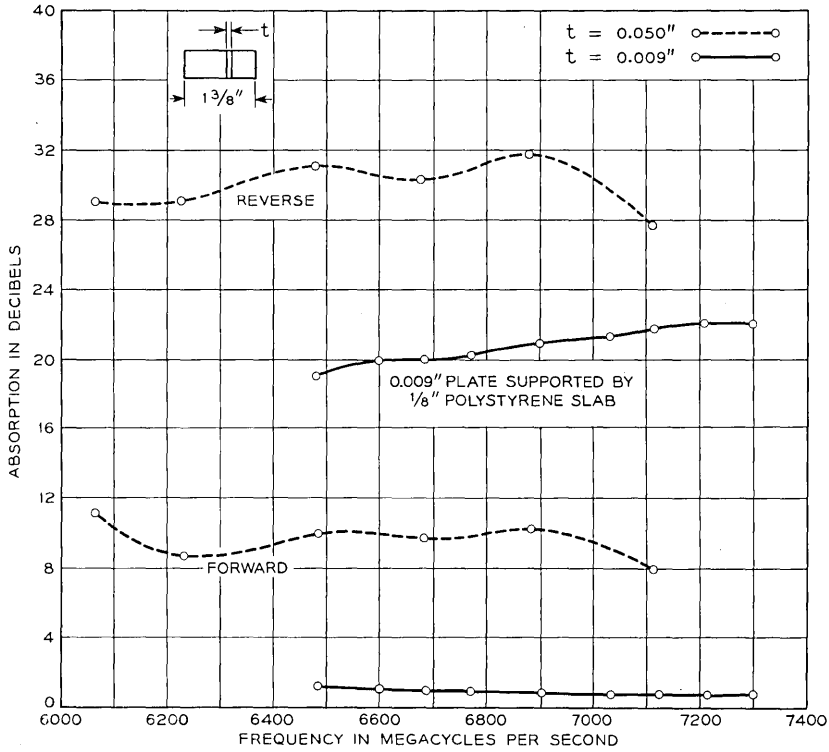


FIG. 21 — Absorption of forward and reverse waves at the optimum value of field as functions of frequency. Two sample thickness are shown to illustrate the effect of sample thickness.

having a loss characteristic such as is shown in Fig. 5, Curve C, in which the loss differential becomes great as soon as the material is saturated, then the bandwidth would be as shown in Fig. 21. Of all the materials which have been measured in the past two years only one sample has shown this effect. At the time such behavior was regarded as the antithesis of the ideal, and no further investigation was made. Now that there is a use for such a material, effort is being directed toward maximizing the effect. Since the effect is thought quite definitely to be due to a very high effective anisotropy field arising either from crystalline anisotropy or demagnetizing effects, the problem of creating the proper material should be quite straightforward.

At this point we should consider the relative merits of the transverse field non-reciprocal devices and those employing the Faraday rotation. We have seen that it is possible to construct a simple isolator using the non-reciprocal absorption in rectangular waveguide. Such an isolator has a minimum forward loss of more than 0.5 db for 20 db reverse loss where an isolator employing the Faraday Rotation can be made with less than 0.1 db forward loss for 30 db return loss. On the other hand the transverse field isolator is much simpler, more compact and easier to

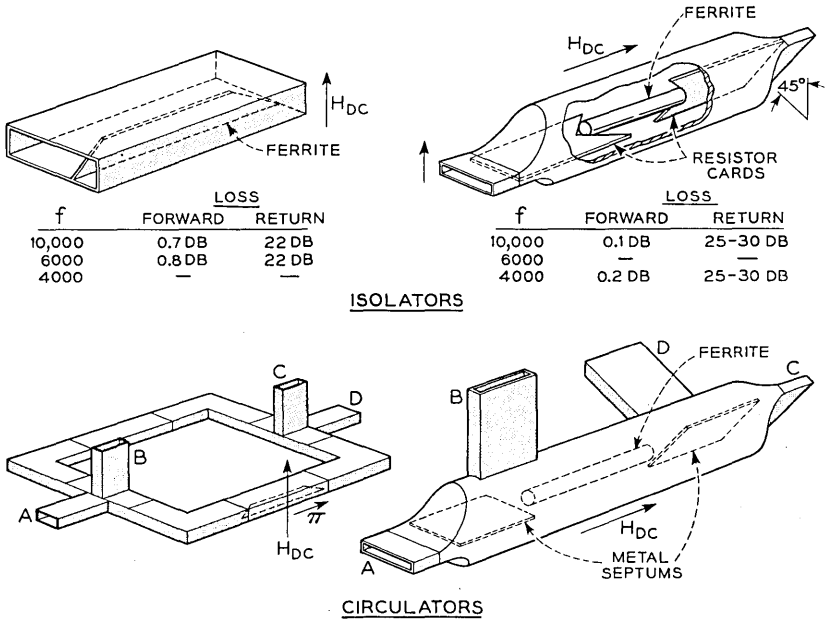


FIG. 22 — Comparison of two basic non-reciprocal elements of the Faraday effect and transverse field effect types.

match impedance-wise. To illustrate the comparison between comparable devices using each principle we have prepared a figure showing the relative structural complexity and performance standards of each. This comparison is shown in Fig. 22. In general the transverse field devices which depend on differential phase shift have about the same insertion loss as those employing the Faraday effect, while the transverse field devices which depend upon differential absorption are somewhat more lossy but simpler to construct than the corresponding Faraday Rotation devices.

SUMMARY

This paper has reviewed the plane wave theory, extended it to discuss waveguide effects, analyzed the various loss mechanisms present in ferrites at microwaves, and discussed numerous measurement techniques and results. It is known that there are original papers in preparation in Bell Telephone Laboratories, and possibly elsewhere, which make this review incomplete at the time of writing. However, it is hoped that the information summarized herein will be of assistance to those who are seeking orientation in this new and rapidly expanding field.

In conclusion the author wishes to thank A. G. Fox, M. T. Weiss, J. P. Schafer, H. Suhl, A. D. Perry and L. R. Walker for permission to discuss herein some of their work and ideas which have not previously been published. The cooperation of F. J. Schnettler and L. G. Van Uitert in providing us with a wide variety of excellent ferrites and in suggesting useful variations in ferrite properties has been of tremendous help in the advancement of the art. The advice given by C. L. Hogan, J. K. Galt and H. Suhl in discussions has proved invaluable and finally the author wishes to thank J. L. Davis for the able assistance he rendered.

Cold Cathode Tubes for Transmission of Audio Frequency Signals

By M. A. TOWNSEND and W. A. DEPP

(Manuscript received August 13, 1953)

Cold cathode gas filled tubes have been extensively applied as electronic switching elements in the telephone system. In general, these applications have been limited to control circuits. The usefulness of these tubes can be further extended by making them capable of carrying voice frequency signals. The transmission properties that are required of the tube for this use are considered. It is shown that troublesome oscillatory noise can be eliminated and that the insertion loss of the tube can be reduced to a low value. Furthermore, by a special design of cathode a stable insertion gain of a few db may be realized. Other requirements on bandwidth, power and distortion are satisfactorily met. Thus, these tubes are potentially useful in coordinate type switches in which voice frequency signals must be rapidly switched.

INTRODUCTION

Cold cathode glow discharge tubes have found increasing usefulness as two-valued switching elements in telephone and other automatic control systems. In many applications the tube functions as a simple control element which either does or does not pass current, depending on the control signals applied. The output signal is in the form of a voltage or a current which can be used to trigger other tubes or operate relays.[1]

In other types of circuits the glow discharge tube may be used as a switch in series with a transmission path for audio frequency signals. When used for this purpose, the tube not only must fulfill the switching requirements but also must meet an additional set of requirements which may be realized by controlling the dynamic properties of the discharge. This paper describes some of the characteristics of cold cathode tubes which have been developed for use as switching elements in series with voice frequency telephone transmission circuits.

TRANSMISSION REQUIREMENTS OF AN ELECTRONIC SWITCH

When an electronic switch is used as a substitute for a pair of metallic contacts, a number of requirements must be met in order that voice

frequency currents be faithfully transmitted. These are as follows:

1. *Gain or Loss* — The impedance level of telephone voice frequency circuits is usually of the order of several hundred ohms. The impedance of any device inserted in this circuit should be sufficiently small compared to this value in order to restrict the insertion loss to less than 1 or 2 db. Of course, the impedance levels of the circuits may be raised to higher values. Aside from the extra cost of transformers, the problems of noise pickup and crosstalk become bothersome when this is done. Since no amplification of the signal from the telephone set is normally used in local transmission circuits, any value of loss is highly undesirable. In fact, since the use of the electronic switch may require the introduction of other circuit elements which introduce loss, it would be desirable that the electronic switch provide a small amount of gain.

The above discussion has considered the gain or loss of the switch in its "closed" condition. Between two busy circuits, there may exist paths consisting of one or more "open" switches. Each of these paths may contribute crosstalk into the circuits. Therefore, impedance required of an individual switch must be high. For a large coordinate switch, there will be a very large number of these undesired paths. This requires that the open impedance of the individual switch be of the order of hundreds of megohms.

2. *Bandwidth* — It is desirable that the electronic switch transmit faithfully frequencies of 300 to 3,500 cycles per second.

3. *Power Output and Distortion* — Since the impedance of the electronic switch may vary with the current passing through it, distortion may be introduced when the current swings are large. On the other hand, without the proper current swing, insufficient power will be delivered to the load impedance. Telephone circuits need to handle powers of the order of a few milliwatts with a harmonic distortion less than a few per cent.

4. *Noise* — The noise introduced into a telephone switching system by an electronic switch should not be noticeable to a subscriber. This means it should be below 10^{-8} or 10^{-9} watts.

5. *Stability* — The properties that have been considered above must be highly stable with time. In central office use, such devices might be used for periods of ten to forty years.

STATIC CHARACTERISTICS

A common form of glow discharge tube comprises a pair of metal electrodes in a glass envelope which is carefully evacuated and filled with

a chemically inert gas to a pressure ranging from 1 to 100 mm of mercury. The negative electrode, called the cathode, is given a special processing which permits it to emit electrons readily when bombarded with positive ions of the gas. The positive electrode, the anode, serves to collect electrons emitted by the cathode as well as those produced in the gas by ionization.

A typical voltage-current characteristic is shown in Fig. 1. For discussion purposes we may divide the curve into several current ranges. In current range I a small residual current flows even at low voltages because of ionization resulting from cosmic rays or radio-active material placed in the tube. The two curves in range I are for different residual currents. At higher voltages this residual current is amplified as a result of additional ions and electrons formed in the gas but it is still extremely small. If the tube voltage is increased still more, the current increases very rapidly until in current range II, a self-sustaining discharge is established. Each electron released from the cathode gains enough energy on the way to the anode so that it produces a large number of positive ions, excited atoms, and photons in the gas. When these particles, arriving back at the cathode, on the average, release another electron, the

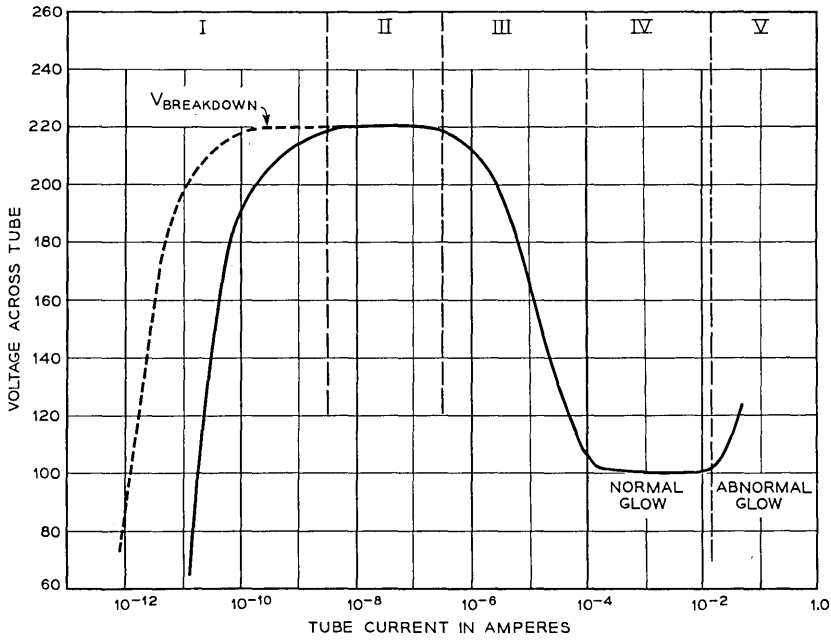


Fig. 1 — Volt-ampere characteristic of cold cathode diode.

self-sustaining condition has been achieved. The value of voltage at which the discharge becomes self-sustaining is usually referred to as the breakdown voltage. A detailed discussion of current ranges I and II has been presented by Druyvesteyn and Penning.[3]

Because the tube is a good insulator in current range I at low voltage and suddenly becomes a good conductor in range II, we often think of a gas tube as a voltage controlled device and use it as such in switching circuits. Actually, however, in current range II the tube can be controlled only by regulating the current. For this reason in the remainder of this paper, the current is considered as the independent variable.

In current range III the voltage falls rapidly with increasing current. A space charge of positive ions begins to develop close to the cathode. By the beginning of current range IV most of the voltage drop in the tube appears across this space charge layer. This region of nearly constant voltage with increasing current is called the normal glow discharge. The space charge layer immediately in front of the cathode is commonly called the cathode dark space. Electrons emitted from the cathode as a result of positive ion bombardment and other processes are accelerated through the high field of this cathode dark space and produce an adjoining layer of intense ionization and excitation called the negative glow. In tubes of the type considered here, this negative glow is the most luminous part of the tube. Beyond the negative glow toward the anode is the so-called Faraday dark space in which no new ionization or excitations are produced. Electrons from the negative glow can readily flow through this region to the anode because their space charge is almost completely cancelled by positive ions diffusing from the negative glow.

Over current range IV the cathode current density is nearly constant. This means that the cross-sectional area of the discharge increases in proportion to the current. This is evidenced by the familiar spreading of the negative glow with increasing current until it covers the entire cathode area.

In current range V the cathode is completely covered with the negative glow and the current density must increase in direct proportion to the total current. This range of currents is called the abnormal glow discharge.

The current-voltage characteristic of a cold cathode for current ranges IV and V may be modified by changing the geometry from a single plane cathode to a hollow cathode.[3] A hollow cathode is one in which there is an overlapping of the regions of cathode fall and negative glow from two portions of the cathode. This overlapping can occur on the inside of a cylindrical or spherical surface or between two plane cathodes more or

less parallel to each other and closely spaced. The optimum dimensions are a function of the density and kind of filling gas used. Electrons and ions generated in the cathode fall and negative glow regions of one cathode can be more efficiently used in producing new electrons and ions if they can enter another cathode fall region instead of diffusing outward into the Faraday dark space. This effect is particularly noticeable in the abnormal glow range of currents.

This is illustrated by the curves of Fig. 2 which were taken on a tube containing two parallel plane cathodes with an adjustable spacing between them. Because Fig. 2 is plotted to a linear current scale, current ranges I and II of Fig. 1 are compressed to the left-hand axis. The upper curve is obtained when the cathodes are far apart so that there is no interaction. The lower curve is obtained when the spacing is close enough to give a hollow cathode effect.

The sustaining voltage of a normal glow discharge is dependent upon the anode to cathode distance. To investigate this we can arrange a parallel plane cathode-anode structure with the anode attached to a piece

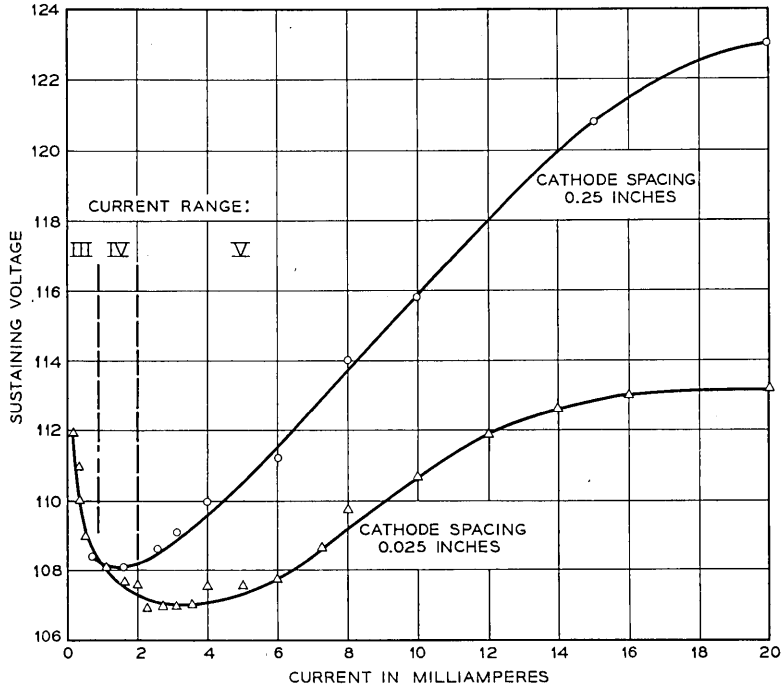


Fig. 2 — Volt-ampere characteristic of parallel plane cathodes at two different spacing.

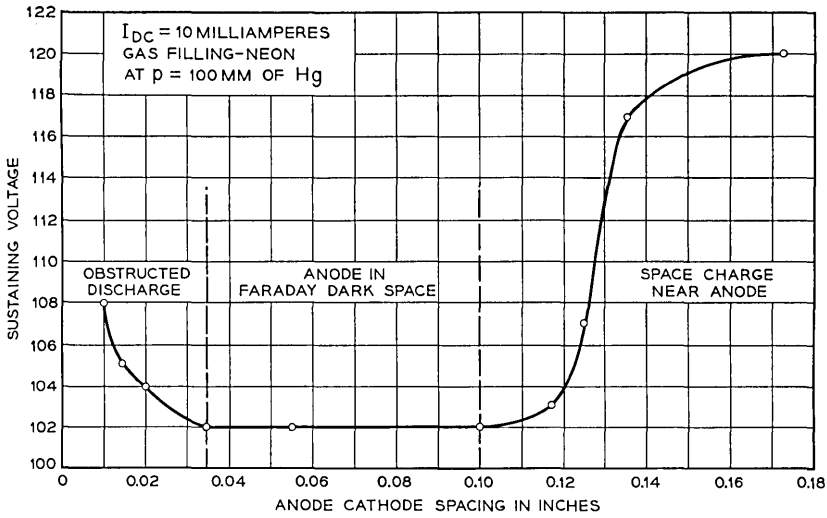


Fig. 3 — Sustaining voltage as a function of anode-cathode spacing.

of magnetic material which can be moved by a magnet external to the tube. Fig. 3 shows a typical set of data obtained in this manner with an operating current in the normal glow range of current. At very close spacings of the order of the negative glow distance, the voltage is higher than normal because some of the electrons released from the cathode are able to strike the anode before dissipating their energy in producing ions and excited particles in the gas. This is referred to as an "obstructed discharge".[3] The voltage must, therefore, be higher because electrons which do produce excitation or ionization must be given extra energy in order to maintain the current. As the anode is moved away from the cathode through the Faraday dark space, the tube sustaining voltage stays at a fairly constant minimum value. This is possible because at close spacings more than a sufficient number of electrons and ions are present in the Faraday dark space to carry the current. As the anode is moved away the Faraday dark space is lengthened. Ions and electrons needed to carry the current diffuse from the cathode region. At sufficiently large distances, however, the ionization density decreases so that not enough ions are present to cancel space charge near the anode. A space charge sheath builds up in the anode region and the sustaining voltage rises with increasing distance.

When the voltage has increased by about 10 to 18 volts depending upon the gas filling, it begins to level off with increase in distance. At about this point, an anode glow may appear in front of the anode. This is due

to the fact that some of the electrons have gained suffcient energy for excitation. A slight further increase in anode-cathode distance usually results in the anode glow changing from a uniform layer to a "ball" of glow. When this occurs, oscillations of several volts amplitude appear across the tube terminals. These oscillations result from a sequence of events which is initiated when the electrons gain enough energy in passing through the anode fall region so that they may ionize. A small number of ions generated in this region will, because of their relatively low velocity, enable a large electron current to flow without developing space charge. This then will reduce the voltage appearing across the anode fall and greatly reduce the number of ionizations taking place. As soon as the recently produced ions leave this region the voltage drop across this region increases causing the ionization to build up again. This alternate building up and decaying of ion density results in the observed oscillation which is ordinarily in the frequency range from 0.5 to 20 kilocycles per second.

This oscillation usually cannot be prevented by external circuit means. However, by proper choice of anode-cathode spacing, type and density of the gas filling, and to lesser extent the geometry of the anode, a tube can be made which is free from anode oscillations. The main restriction that this puts on tube design is the limitation of breakdown voltage.

IMPEDANCE

From the previous discussion of transmission requirements it is clear that one of the most important properties of a gas tube is the impedance presented to small ac signal currents superimposed on the steady dc operating current. At low frequencies, these signals cause the voltage across the tube to vary in accordance with the static characteristic. The impedance of the tube to these signals is almost entirely resistive and is equal to the slope of the static characteristic. At higher frequencies, however, there is a lag in the adjustment of the voltage across the discharge to the changes in current. Hence, at these frequencies, the impedance of the tube may have both resistive and reactive components. This is illustrated in Fig. 4.

The small superimposed signals result in current-voltage loci which are ellipses. In current range III at 200 cps, the position of the ellipse corresponds to a negative resistance in series with an inductance. The negative resistance changes rapidly with frequency and as shown at 2,000 cps, it may be positive. Because of the rapid variation of impedance, both with frequency and current, this range of currents is not generally useful for dependable transmission of voice frequency signals.

In the higher current range IV and V the impedance can be represented by a positive resistance in series with an inductive reactance. The resistance increases slowly with frequency as indicated by the elliptical current-voltage loci at 200 and 2,000 cps.

The impedance of a cold cathode tube also varies with anode-to-cathode spacing. This may be studied by means of the same movable-anode parallel-plane tube used to obtain the data of Fig. 3. We again operate the tube in the current range of the normal glow (IV) and, for illustrative purposes, choose a measuring frequency of 1,000 cycles per second. The results are shown in Fig. 5. The resistive and reactive components of tube impedance are independent of anode spacing throughout the Faraday dark space and well into the obstructed discharge region. At large distances where the electron space charge sheath begins to build up in front of the anode, the impedance increases rapidly with distance.

Some useful conclusions about the design of transmission tubes can be drawn from the data of Fig. 5. We can consider the total tube impedance as being made up of the sum of the impedances introduced by the various regions of the discharge. Since large variations in the length of the Faraday dark space do not affect tube impedance, it is concluded that this region has negligible impedance. This means that, so long as the anode-to-cathode distance is short enough to avoid a space charge sheath at the anode, we can concentrate our attention on the cathode fall region. The following detailed discussions of impedance are consequently restricted to the cathode portions of the glow discharge.

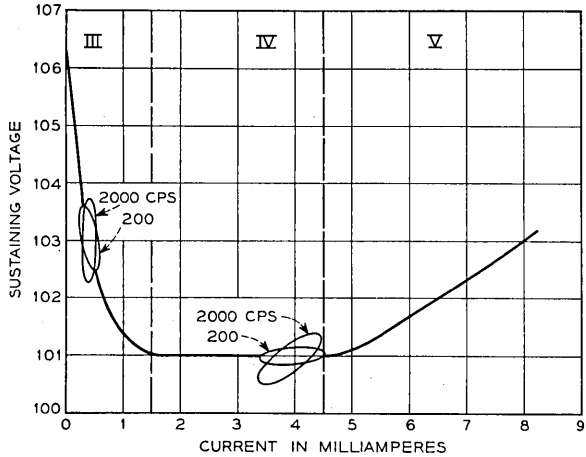


Fig. 4 — Static and dynamic volt-ampere characteristics.

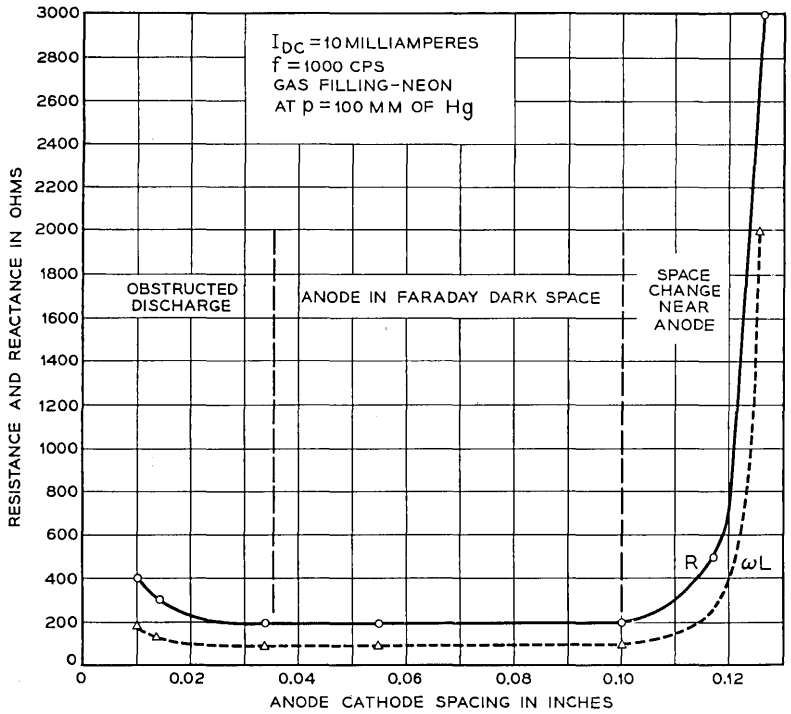


Fig. 5 — Resistive and inductive components of impedance as a function of anode-cathode spacing.

IMPEDANCE OF PLANE CATHODE TUBES

Let us consider first the impedance of a tube with a plane cathode or of a cathode with a radius of curvature large compared to the combined thicknesses of the cathode dark space and the negative glow. An example is the Western Electric 313-C which has a gas filling of 99 per cent neon, 1 per cent argon and a cathode surface of barium and strontium oxides. At a fixed steady current of 25 ma flowing through the starter gap, it is found that the resistive component of the impedance varies with frequency as shown in Fig. 6. In the middle of the voice band it has a value of about 200 ohms. There is also an inductive reactance of about 65 ohms at this frequency. At a fixed frequency, it is found that the resistive component of this type of tube decreases with current approximately inversely as the square root of the current. This is shown in Fig. 7. The

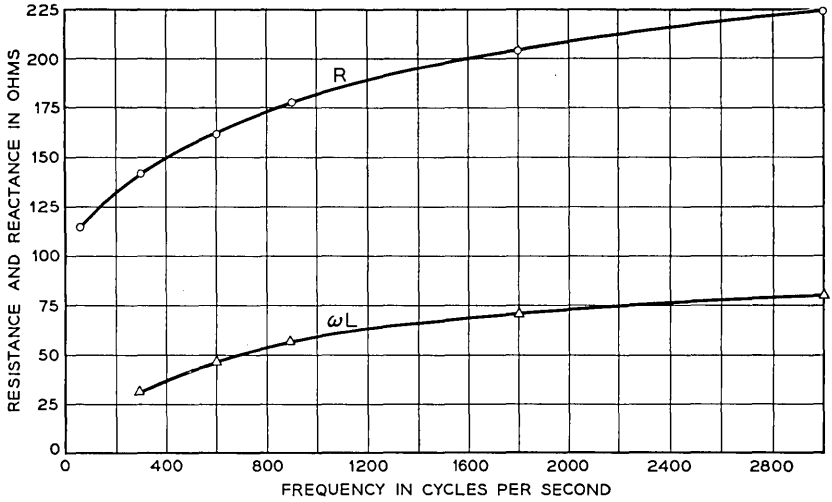


Fig. 6 — Resistive and inductive components of 313-C starter gap impedance as a function of frequency. I_{dc} = 25 mm.

inductive reactance is related to the current approximately as follows:

$$\omega L = B + \frac{C}{I_{dc}}. \tag{1}$$

B and C are functions of frequency.

If we wish to use tubes of this type in central office switching circuits which pass voice frequency currents, we find that even one tube connected between a 600-ohm source and a 600-ohm load gives a prohibitively high insertion loss. The impedance level of the voice frequency circuit may, of course, be raised up to 4,000 or 5,000 ohms. But practical switching circuits might require as many as four tubes in series. Hence, we see that the impedance of tubes of this type severely limits their usefulness.

From Fig. 7 we might argue that to get low resistive components we might continue to increase the direct current flowing through the tube. However, in this range of currents, the life of the tube varies approximately as $1/I_{dc}^{3}$ while the resistive component varies as $1/I_{dc}^{1/2}$.

Over the range of currents plotted in Fig. 6, the cathode is fully covered with glow. As noted above, the resistive component is

$$R = \frac{A}{I^{1/2}}. \tag{2}$$

If instead of passing a given direct current through a single tube we pass the same current through a parallel combination of n tubes, the resistive component of the combination would be

$$R \text{ comb.} = \frac{1}{n} \frac{A}{\left(\frac{I_{dc}}{n}\right)^{1/2}}, \qquad (3)$$

$$= \frac{1}{n^{1/2}} R .$$

The resistive component of the combination is $\frac{1}{n^{1/2}}$ times that of a single tube passing the same total current. This assumes, of course, that the cathodes of the tubes are covered with glow. Obviously this is not a practical method of attaining a low impedance because of the instability of paralleled individual tubes. It does suggest, however, that by increasing the cathode area of a single tube until the glow just covers the full area, a lower impedance is obtained. This has been done experimentally.

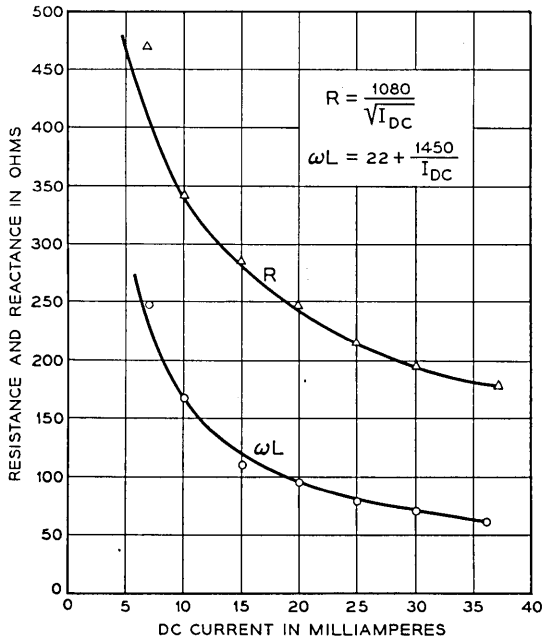


Fig. 7 — Resistive and inductive components of 313-C starter gap impedance as a function of direct current.

Fig. 8 shows the data for a tube of this type which has a cathode area about six times that of the 313-C. This change in cathode area along with the necessary change in anode geometry led to an impedance reduction of about five times. This type of design, of course, gives a large increase in life of the tube but it is undesirable from the standpoint of tube size.

As pointed out above, low impedance can be attained by an increase of the tube current and within limits by an increase of the cathode area. A third parameter which may be varied is the density of the gas filling. At a constant current and with a given cathode area, the resistive component of the impedance decreases with increase in density or the pressure, p, at a fixed temperature approximately in accordance with the relation

$$R\alpha \frac{1}{p^2}. \tag{4}$$

The curve in Fig. 9 was obtained over a limited range of argon fillings with a barium strontium oxide cathode. Since the current density increases approximately in proportion to p^2, the effective cathode area will tend to be reduced unless the total current is kept high enough to cover the cathode fully. Therefore, with a fixed total current there is a limit to either increase in pressure or in cathode area.

In further search for low impedance, a number of different gas fillings have been tried. They have included all the rare gases as well as several mixtures of them. No significant advantages were obtained by the use of other than the more common neon or argon gas fillings.

We therefore see that the ability to control impedance properties of

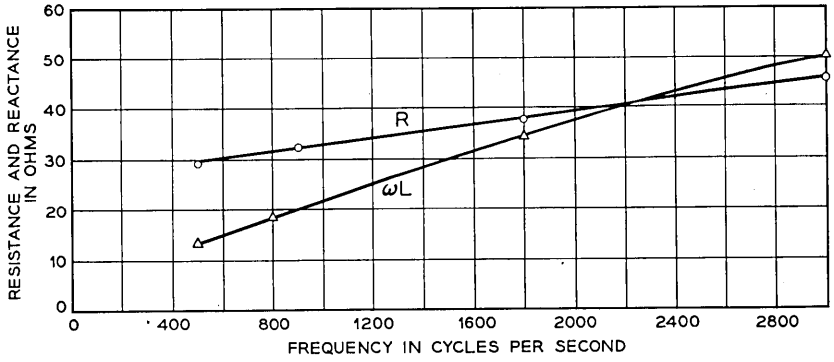


Fig. 8 — Resistive and inductive components of impedance for a large area cathode tube as a function of frequency. I_{dc} = 25 mm.

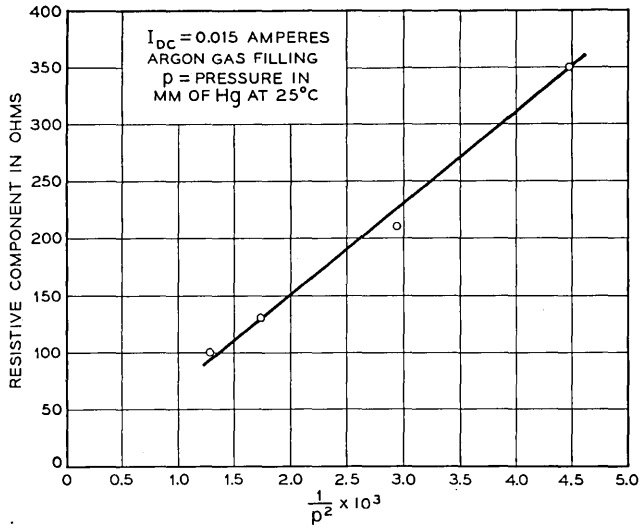


Fig. 9 — Resistive component of impedance as a function of density (pressure at 25°C) of gas filling.

the plane cathode through changes in tube current, cathode area, density and type of gas filling is limited. A reduction by a factor of five or ten times over the values present in commercial tubes is certainly possible. But the possibility of obtaining values of less than +10 ohms or negative values seemed remote. Consequently, development effort was eoncentrated on the hollow cathode tube described below.

HOLLOW CATHODE TUBES

The static voltage-current characteristic of a hollow cathode was shown in Fig. 2 to be below that of a plane cathode of the same area when operating at currents in the abnormal glow region. It has been found that by proper choice of the cathode dimensions and the kind and density of the filling gas, desirable transmission characteristics can be achieved. The following discussion illustrates the manner in which some of the variables are interrelated.

We will consider a "U" shaped cathode which has been formed by folding a piece of molybdenum sheet in the form illustrated in Fig. 10. The choice of dimensions of the hollow portion will be discussed later. The anode is a cylindrical rod placed in front of the cathode. The anode-to-cathode spacing is selected so that the anode is always within the Faraday dark space and hence does not influence the impedance appreciably. It is assumed that the structure of Fig. 10 is sealed in a bulb

which has been carefully evacuated and filled to a fraction of an atmosphere with a rare gas such as neon.

A typical static volt-ampere characteristic of the structure of Fig. 10 is shown in Fig. 11. This curve was taken with a cathode having a hollow portion $\frac{1}{8}''$ long and $\frac{1}{16}''$ deep with a cathode gap of 0.023″. A neon filling pressure of 58 mm of mercury was used. It can be seen that there is the usual low-current negative slope associated with the transition from breakdown (current range III of Fig. 1). A new characteristic of interest is the second region of negative slope in the abnormal glow range of currents. It has been found that this second region can be made stable with time in a given tube and reproducible from tube to tube. It is also

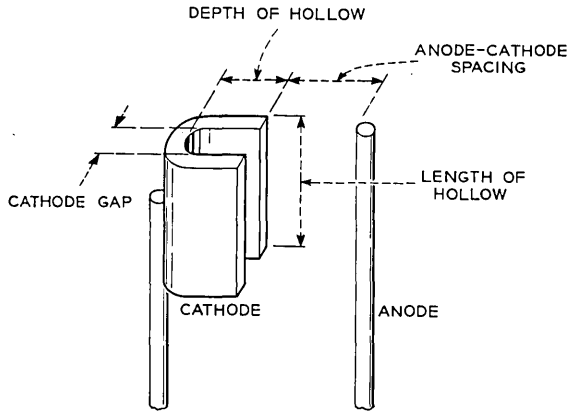


Fig. 10 — Electrode geometry of a hollow cathode tube.

found that the tube impedance has a negative resistance component over the voice frequency range. Thus, this second region of negative resistance offers attractive possibilities as a transmission element.

The impedance of this tube at 300 and 3,000 cps is shown in Fig. 12 for the same range of operating current as Fig. 11. The optimum current for negative resistance and the value of negative resistance are functions of the cathode gap, but so long as the other cathode dimensions are constant the optimum current is relatively independent of the density of the filling gas.

A useful way of studying the interrelation of cathode gap and filling pressure is shown by Figs. 13 and 14. For these data the length and depth of the hollow portion were kept constant at the values of $\frac{1}{8}''$ and $\frac{1}{16}''$ respectively. Fig. 13 shows the resistive component of impedance as a function of frequency for different filling pressures with a fixed cathode gap.

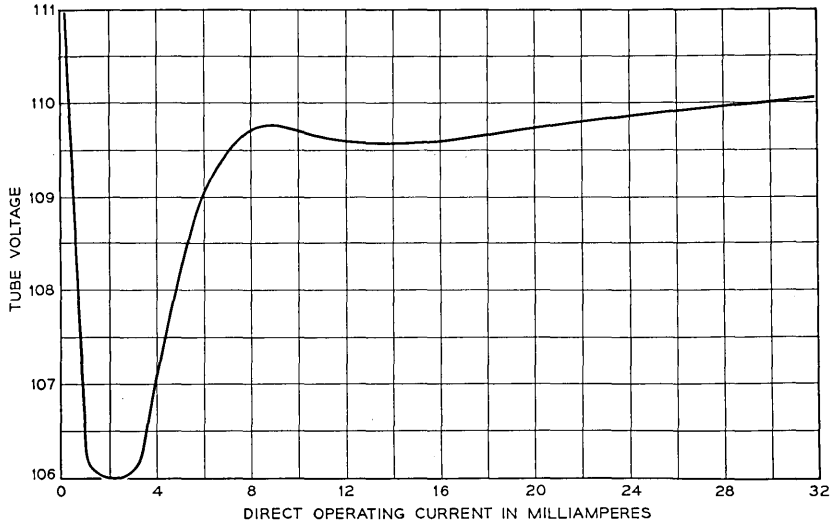


Fig. 11 — Static volt-ampere characteristic of hollow cathode tube.

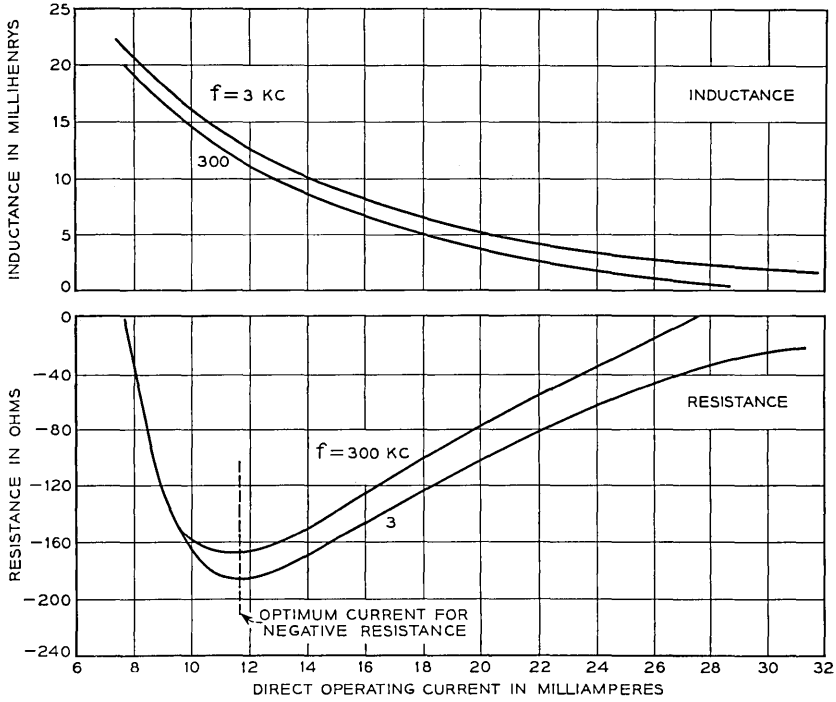


Fig. 12 — Resistive and inductive components of impedance of a hollow cathode tube.

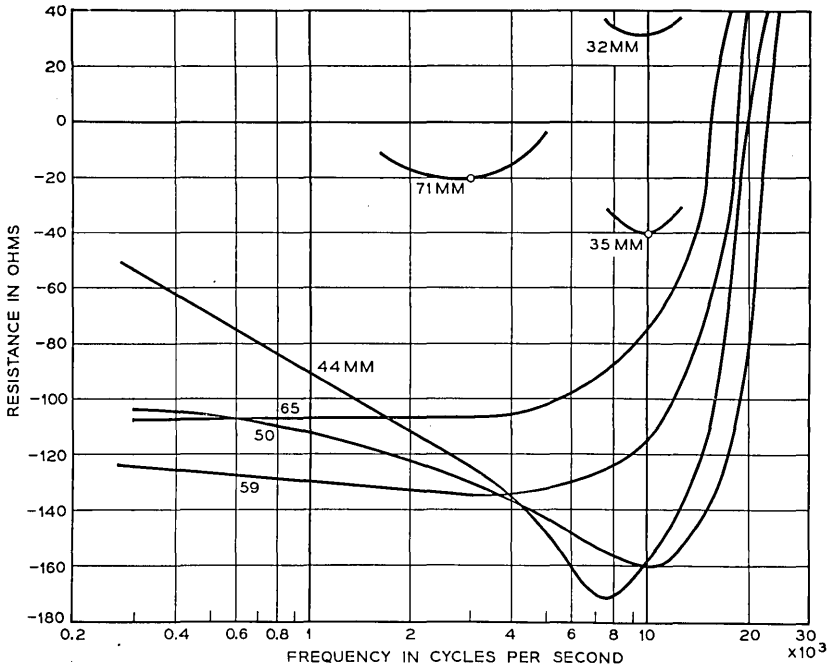


Fig. 13 — Resistive component of impedance as a function of frequency for different neon filling pressures. Cathode gap = 0.024 in.

Fig. 14 shows the variation of the resistive component of tube impedance as a function of frequency for several cathode gaps and at a fixed density of filling gas. For both Figs. 13 and 14, the current was adjusted for optimum negative resistance.

It can be seen from Fig. 13 that the choice of a neon filling gas at a pressure near 60 mm and from Fig. 14 that a cathode gap near 0.024 inch could be expected to yield a negative resistive component of impedance which is reasonably insensitive to filling pressure and which is also constant in value over the voice frequency range. This justifies the choice of cathode gap and filling pressure used in the tube on which the data of Figs. 11 and 12 were taken.

One other parameter of interest is the limit of anode-to-cathode distance. This too is a function of cathode gap and gas density. A typical curve taken for the same cathode geometry and gas filling as used for Figs. 11 and 12 is shown in Fig. 15. It is seen that the negative resistive component is essentially independent of anode distance for a distance of approximately 0.050 inch. This means that the breakdown voltage of

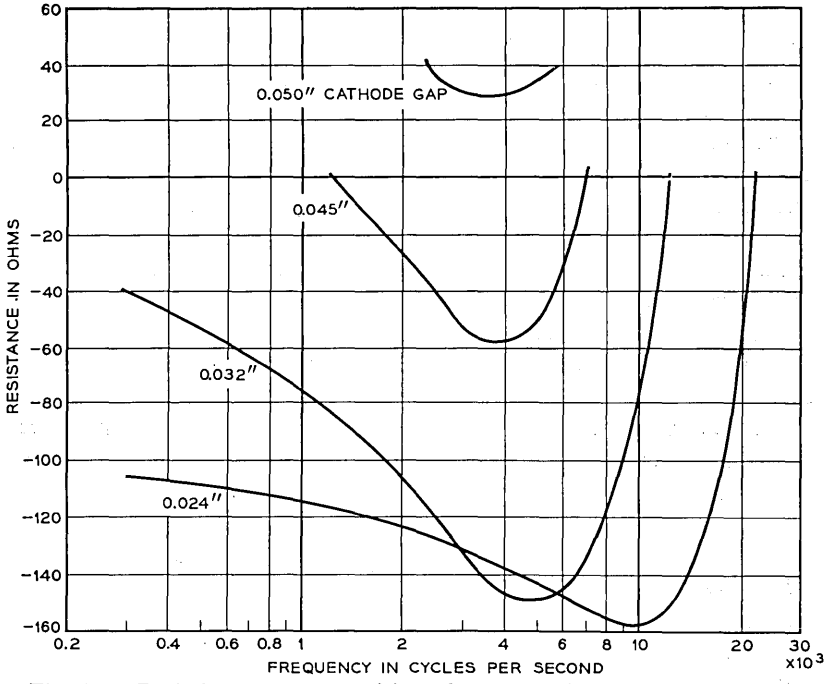


Fig. 14 — Resistive component of impedance as a function of frequency for different values of cathode gap. Neon filling pressure = 50 mm of Hg.

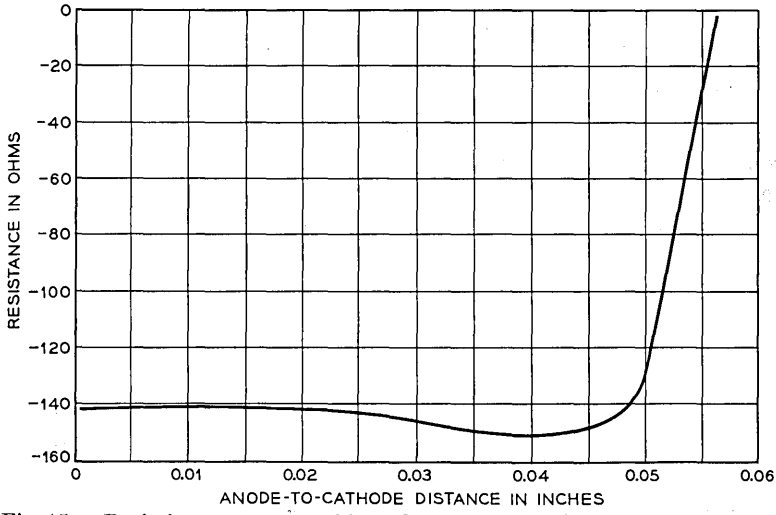


Fig. 15 — Resistive component of impedance as a function of anode-to-cathode distance.

a switching tube can be adjusted by changing this spacing so long as the upper limit is not exceeded.

Although the above discussion is limited to only one adjustable cathode dimension, the cathode gap, other variations are of course possible. The length and depth of the hollow should be at least a few times the width of the cathode gap in order that an efficient hollow be formed. Increasing the hollow length or depth requires larger currents and, if carried too far, the glow may not completely fill the hollow at the optimum current for negative resistance. This is undesirable because the unused portion of the cathode may change its properties with time and produce unstable characteristics. The entire geometry may be scaled to a larger size if the density of the filling gas is reduced by approximately the same factor.

The choice of cathode material is restricted by the high current densities of hollow cathodes. Coatings of alkaline earth oxides or similar materials have too short a life. Pure molybdenum has found to give a satisfactorily low sustaining voltage together with long life and stable operating characteristics. Life tests have shown that tubes can be made which will operate satisfactorily for the equivalent of 20 to 40 years in central office service.

It is seen from the above that by changing the cathode geometry and density of the filling gas a variety of impedance properties can be obtained. The final choice must be determined by the overall transmission requirements. As an example, the transmission performance of a typical tube will now be discussed.

TRANSMISSION PERFORMANCE OF A TYPICAL NEGATIVE RESISTANCE DIODE

The circuit of Fig. 16(a) shows a cold cathode switching tube in series with a transmission path. The voltage across the load resistor R_L under conditions of Fig. 16(a), divided by the voltage across R_L with no tube in the circuit, Fig. 16(b), is one measure of the transmission performance. This ratio, called the insertion voltage gain, is given by

$$\text{I.V.G} = \frac{R_S + R_L}{R_S + R_L + R_t + j\omega L_t} . \tag{5}$$

The derivation of this expression assumes that the transformers are ideal and that the reactance of the condenser C is negligible. Maximum gain occurs at low frequency where the reactive component of tube impedance, $j\omega L_t$, can be neglected. If the resistive component of tube impedance, R_t, is negative, the gain will be greater than unity. The gain approaches an infinite value as the unstable condition is approached where the

negative resistance of the tube equals the sum of the source and load resistance. Large values of gain are not practical because this imposes undesirable restrictions on the constancy of the circuit and tube impedances.

Additional restrictions on gain arise from bandwidth and distortion considerations. If it is assumed that both R_t and L_t are independent of frequency over the voice band, it is possible to use the above equation for an approximate calculation of bandwidth. The half-power point occurs at the frequency f_c at which the reactive term equals the sum of the other three terms in the denominator, or

$$f_c = \frac{R_S + R_L + R_t}{2\pi L_t} \tag{6}$$

Since the gain does not fall off at low frequency, the upper cut-off frequency f_c is a measure of the bandwidth. Substituting R_S and R_L from equation (5) into equation (6) gives

$$f_c\ (1 - \text{Low frequency I.V.G.}) = \frac{R_t}{2\pi L_t}\,. \tag{7}$$

Thus for a given tube an increase in gain is accompanied by a decrease in bandwidth.

As shown in Fig. 12 the impedance of a negative resistance tube is dependent on the current passing through it. This will cause some distortion as the signal current swings above and below the average direct current value. The distortion is small so long as the non-linear tube resistance is small compared to the total circuit impedance.

It can be seen from the above discussion that for a given tube the gain, bandwidth, and distortion are all dependent on the source and load impedances. Fig. 17 shows the experimental performance of a typical tube for the special case where source and load impedances were equal. Insertion gain has been converted to power gain in db. The distortion

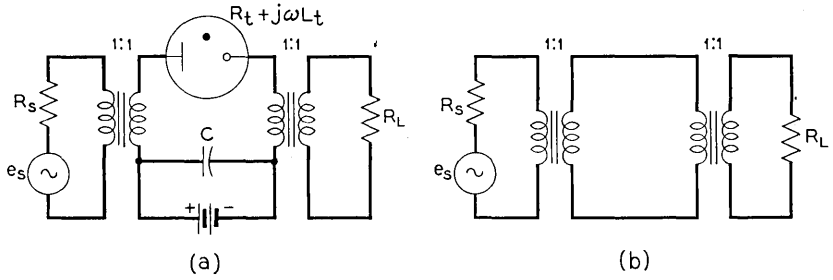


Fig. 16 — Transmission circuits with and without tube.

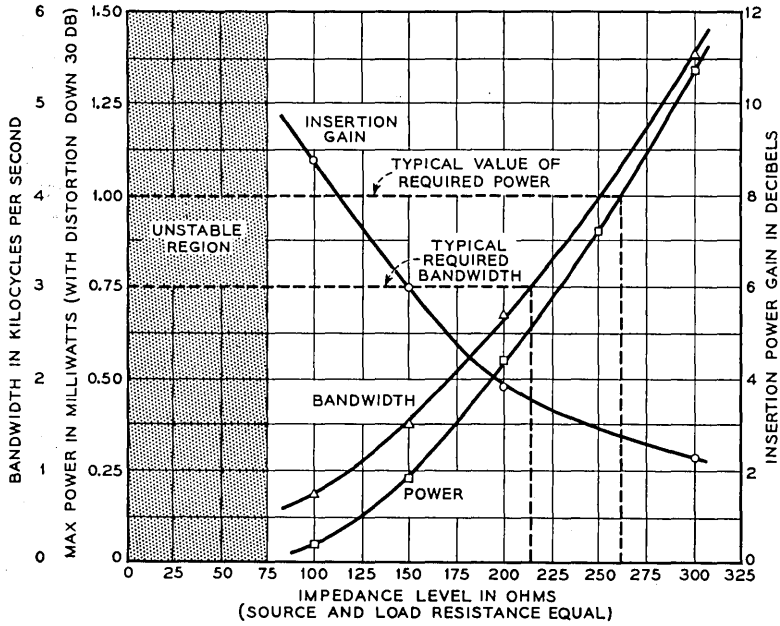


Fig. 17 — Performance curves for experimental cold cathode negative resistance diode.

has been related to the power handling capacity by measuring at each value of source and load impedance, the maximum power output at which the total harmonic distortion is 30 db below the signal.

The noise introduced by a tube affects its usefulness as a transmission element. By designing the tube so that the anode is in the Faraday dark space, anode oscillations are avoided. The remaining noise is at a low level, typical values being 10 decibels above the noise reference level of 10^{-12} watts.

CONCLUSIONS

Cold cathode glow discharge tubes can be made with stable and reproducible impedance characteristics. By proper choice of anode-cathode spacing and pressure of filling gas, it is possible to eliminate oscillation noise associated with the anode region. By properly choosing the density of filling gas and the area of a plane cathode, it is possible to obtain a low positive resistance component of impedance. By proper correlation of cathode geometry and filling gas density, hollow cathodes can be used to obtain a negative resistance component of impedance. Bandwidth, power, distortion, and noise requirements of voice frequency transmis-

sion circuits can be satisfied without sacrificing switching characteristics. Cold cathode glow discharge tubes are, therefore, a potentially useful electronic switch for use in series with voice frequency transmission circuits.

ACKNOWLEDGMENT

The authors wish to acknowledge the contributions of F. T. Andrews in analyzing and measuring the amplifier performance of negative resistance gas tube diodes.

REFERENCES

1. Ingram, S. B., Elec. Eng. (A.I.E.E. Transactions), **58**, pp. 342–346, 1939.
2. Depp, W. A., and W. H. T. Holden, Elec. Mfg., **44**, pp. 92–97, 1949.
3. Druyvesteyn, M. J. and F. M. Penning, Revs. Mod. Phys., **12**, pp. 87–174, 1940.

Balanced Polar Mercury Contact Relay

By J. T. L. BROWN and C. E. POLLARD

(Manuscript received June 19, 1953)

A new type of relay making use of solid contacts maintained continuously wet with mercury has been developed. It has a symmetrical polar structure, resulting in improved sensitivity and speed compared with the existing neutral structure with similar contacts. It is also particularly well adapted to switching of high frequency circuits. Two magnets are used for polarization, and the relay is adjusted after assembly to desired values of sensitivity for operation in both forward and reverse directions by selective adjustment of the magnet strengths.

INTRODUCTION

In a previous paper[1] a mercury contact relay is described in which the contact elements are maintained wet with mercury through a capillary path to mercury reservoir below the contacts. The present paper describes a new design making use of this same mercury contact principle, but with a symmetrical polar structure which gives improvement in sensitivity and speed capabilities over the previous neutral structure.

VERTICAL RELAY

Fig. 1 shows one design of the relay, adapted for general use in a vertical position. It provides a single pole double throw magnetic switch in a sealed glass tube, enclosed along with an operating coil and polarizing magnets in a steel can with a medium octal plug base similar to the previous type relay (275 type).

Fig. 2 shows the glass enclosed magnetic switch element. The armature is a tapered reed welded to a tubular stem which is sealed in the glass at the lower end. Mercury, and gas under pressure are introduced through this tube, which is then welded closed. The magnetic working gaps are formed between the armature and fixed pole-pieces which are sealed in

[1] J. T. L. Brown and C. E. Pollard, Mercury Contact Relays, Elec. Eng., **66**, Nov., 1947.

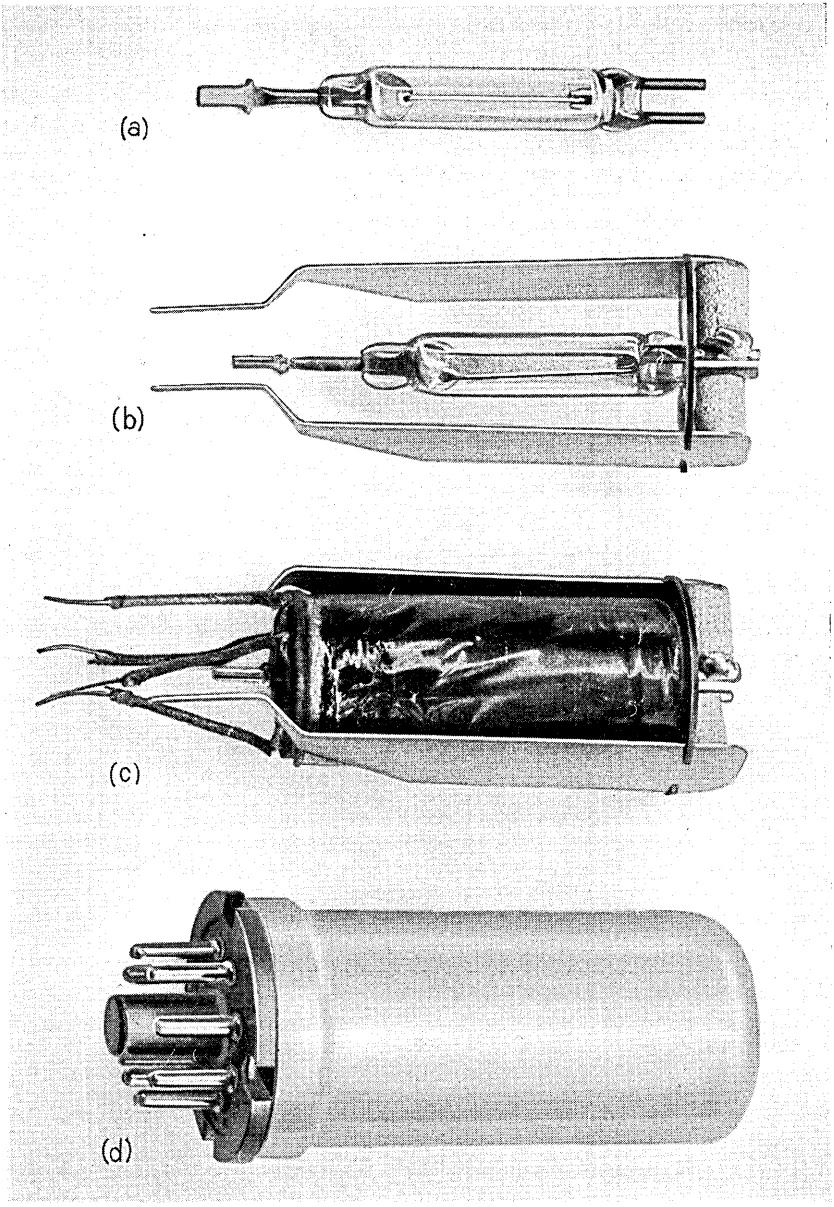


Fig. 1 — (a) Switch element. (b) Magnets, spool, and side plates added. (c) Coil added. (d) The complete relay.

the glass at the upper end. The fixed contacts are made from small platinum alloy balls which are welded to the pole-pieces. The armature strikes the fixed contacts at a point which is close to a node for one of its principal modes of vibration. This limits the bounce on impact to an amplitude which does not open the mercury bridge which is formed.

Mercury in a pool at the bottom of the switch is fed to the contact area through grooves rolled in the armature surface. Except for the platinum alloy contacts, the surfaces of the pole-pieces inside the switch have an oxide coating which prevents them from being wet by the mercury. This limits the mercury bridge which is formed between the armature and pole-piece to the area of the platinum alloy contact. A ceramic detail is inserted between the pole-pieces at the top of the switch. The ceramic is specially resistant to wetting by the mercury and thus prevents mercury from collecting between the pole-pieces.

The polarizing magnets are soldered to the pole-piece terminals outside the switch. Permalloy plates are soldered to the outer poles of the magnets and extend down on the outside of the coil, forming a return path to the lower end of the armature. The coupling at the lower end is made relatively loose in order to limit magnetic soak effects. The steel can provides a magnetic shield.

STATIC MERCURY CONFIGURATION

The static configuration of the mercury surfaces in the switch depends upon the shape and contact angle to mercury of the solid surfaces with which it comes in contact, and the curvature of the free mercury surfaces. This curvature depends on the surface tension of the mercury surface T and the pressure difference Δp between the inside and outside of the mercury at the point under consideration. That is,

$$\Delta p = T\left(\frac{1}{R_1} + \frac{1}{R_2}\right), \tag{1}$$

where R_1 and R_2 are principal radii of curvature of the surface (radii taken in planes cutting the surface at right angles to each other).

In this switch the pressure difference is a function of the height of the point under consideration above the surface of the reservoir:

$$\Delta p = \rho g h. \tag{2}$$

Substituting the values

ρ = 13.6 g/cc for mercury,
g = 980 cm/sec^2, and
T = 450 dynes/cm for mercury,

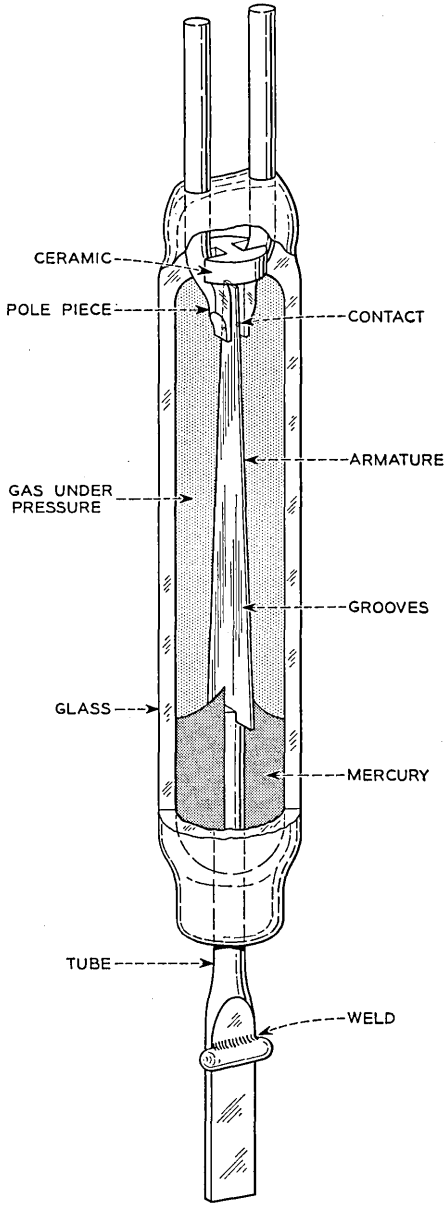


Fig. 2 — Vertical switch.

we obtain

$$h = 0.0338 \left(\frac{1}{R_1} + \frac{1}{R_2} \right) \text{ cm.} \tag{3}$$

Fig. 3 shows a cross section of the armature and pole-piece at the contact, indicating the configuration of the mercury in the grooves and around a closed contact, as determined from equation (3). In the grooves, it will be noted, the surface is concave cylindrical, with a radius of -0.0163 cm, providing a path from the reservoir at this height with nearly the full capacity of the groove. The contact angles to the armature

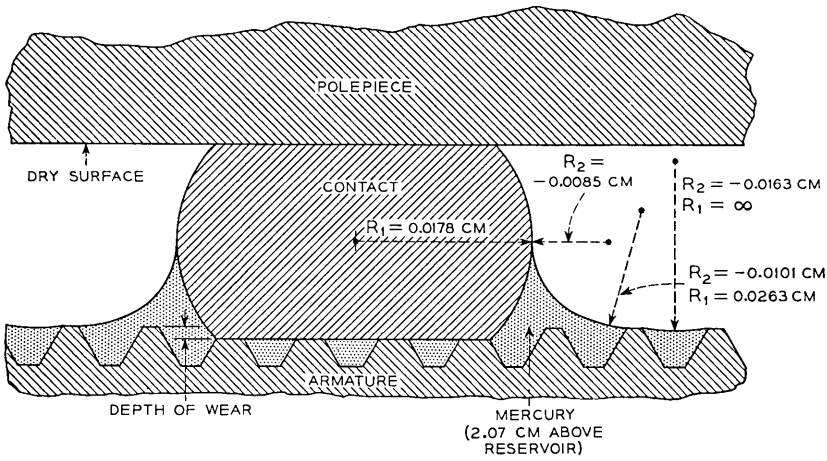


Fig. 3 — Cross-section showing configuration of mercury surface in grooves and around closed contact.

and platinum alloy contact surfaces are all zero, as these show complete wetting. The mercury fillet around the contact has positive curvature in the plane of the armature ($R_1 = +0.0178$ cm) and negative curvature ($R_2 = -0.0085$ cm) in planes through the axis of the contact.

The depth of wear indicated in Fig. 3 is brought to a stable value by an aging process in production.

DYNAMIC MERCURY CONFIGURATION

Fig. 4 shows flash photographs of the contacts at various instants during operation at 60 cps. It will be noted that, as the armature moves away from the fixed contact on the right, Figs. 4(a), (b) and (c), a bridge is drawn out which breaks in two places, leaving a free drop which falls out, Fig. 4(c).

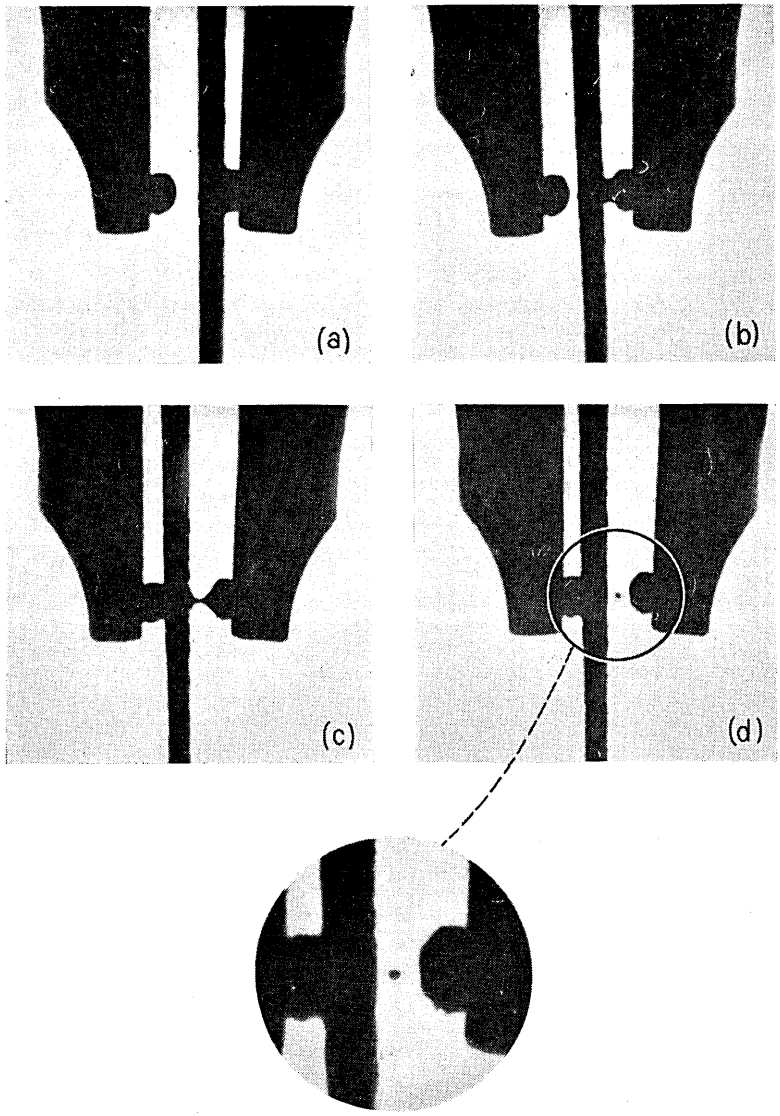


Fig. 4 — Dynamics of mercury contact. (a) Armature against contact on right. Note spherical shape of mercury on left contact. (b) Armature moves to left, drawing out mercury bridge. (c) Armature further to left; bridge about to break. (d) Armature reaches contact on left. Mercury bridge has broken in two places, leaving a tiny ball which later drops away.

When the ball falls out as the result of an operation, the loss of mercury results in a small, temporary constriction of the fillets in the grooves near the contact that was just opened. That is, the curvature of the surfaces of these fillets becomes more negative. As shown by equation (1), this produces a local decrease in liquid pressure. Mercury therefore flows up the armature from the reservoir to restore the normal static balance between surface tension and negative head. This is the fundamental behavior of an ordinary wick. The ball drops into the reservoir unless it happens to hit the armature on the way down. Thus, for repeated operations, there is a continuous circulation of mercury.

As indicated in Fig. 3, the fixed contact is a ball, with a flat surface where contact is made to the armature. In Fig. 4(c), taken directly after the breaking of the mercury bridge, the mercury remaining at the fixed contact on the right has been thrown back on the contact by surface tension forces, laying bare the flat surface. After several flow oscillations that are not shown, it comes to rest with the spherical contour shown by the contact at the left in Fig. 4(a). That is, being disconnected from the reservoir and having a limited wet surface to spread over, it assumes a positive head corresponding to a positive spherical radius about equal to that of the contact. This provides a mercury "cushion" in the form of a segment of a sphere, to which contact is made when the relay operates.

ADJUSTMENT OF SENSITIVITY

For various combinations of MMF's (magnetomotive forces) in the two magnets, various corresponding pairs of sensitivity values exist for operation in the two directions. A theoretical analysis of this relationship is given in the attached appendix. The adjustment of magnet strengths to obtain a specified pair of sensitivity values is made on the completed relay, using two electromagnets placed outside the can opposite the relay magnets. An automatic circuit is provided for this purpose that makes a complete adjustment in about 15 seconds, the time being dependent upon the precision required and the uniformity of the product being adjusted. The procedure used and the basis for it are discussed in the Appendix.

All of the adjustments used are of the type for which the armature moves all the way from one contact to the other when an operate current is applied. This represents the condition for the minimum differential ampere turns obtainable between the two operate values because it makes use of armature momentum.

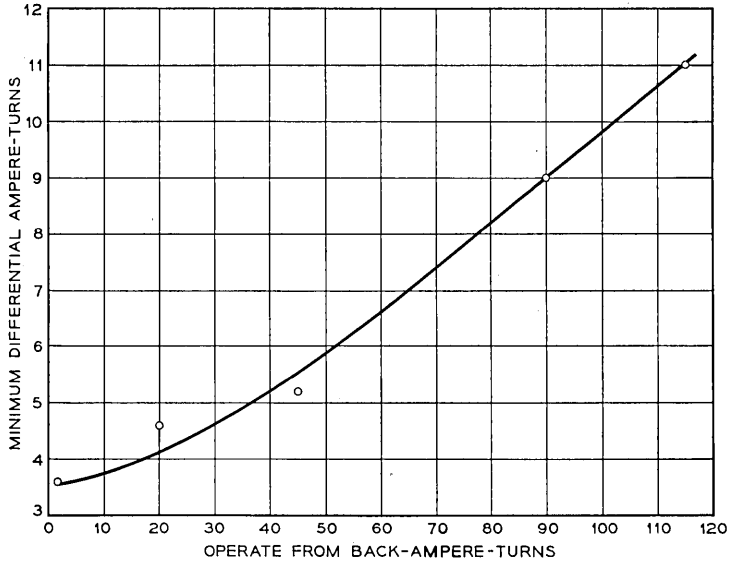


Fig. 5 — Minimum differential ampere turns as function of magnetic bias.

BIAS ADJUSTMENTS

As bias adjustments are made magnetically, rather than with the spring used in some polar designs, the armature flux tends to approach the same value at the operating points as it has in a balanced adjustment with the same differential ampere turn value. Such difference as does enter with increase of bias is due to differences in armature flux distribution between MMF introduced by the coil and that introduced by the magnets.

The effect of bias is illustrated in Fig. 5, which shows the minimum differential ampere turn value obtainable in a relay as a function of the higher operate value. The minimum differential value of 3.6 NI with a balanced adjustment is about the same as is obtainable with other polar relay designs. For operate values of about 100 ampere turns a "release to operate ratio" of about 90 per cent is obtained.

Larger differential ampere turn values, both with and without bias, are of course possible. The amount of bias obtainable in combination with larger differentials is somewhat reduced because of the greater magnet strength required.

EFFECT OF MAGNETIC SOAK ON SENSITIVITY

The effect of "soak" on sensitivity is illustrated in Fig. 6. It shows the change which takes place when the ampere turns to operate are

measured after applying a given ampere turn soak in each of two polarities.

SPEED OF OPERATION

Fig. 7 shows typical operate times of the relay for one balanced adjustment and one biased adjustment. The ordinates correspond to time to effect closure at the opposite contact after input is applied to the coil. The abscissae are shown both in terms of power in the full relay winding and the corresponding number of ampere turns. Two circuit conditions are indicated, one in which the voltage was applied directly to the full winding and the other in which the voltage was applied through a resistance equal to the coil resistance.

The winding used in this case is one which is specially designed for speed rather than sensitivity, being shorter in length, with the working gap near the middle. Its time constant (inductance-resistance ratio) is about 0.0006 second.

Fig. 8 shows "release" time measurements, where the relay, with various biased adjustments, was operated by opening the circuit. The curve is shown plotted against the "operate" setting of the relay. In this form it is relatively independent of the differential ampere turn adjustment.

The time required to open the contact from which the armature moves is typically about 0.0002 second longer than the closure time, as the

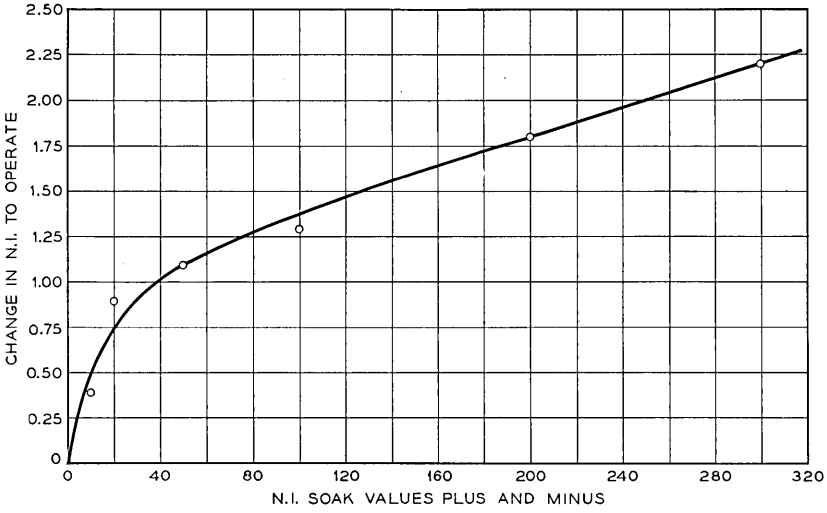


Fig. 6 — Change in NI. To operate versus plus and minus NI soak.

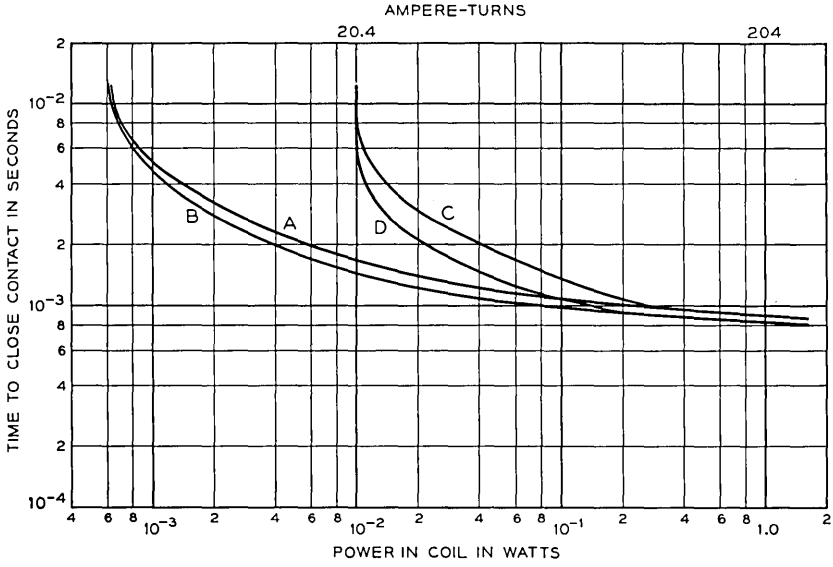


Fig. 7 — Time to close opposite contact after voltage is applied to coil. (A) ±5 NI adjustment, voltage directly on coil. (B) Same as (A), except with resistance equal to coil in series. (C) Operate +20.4 NI, release +9.9 NI voltage directly on coil. (D) Same as (C), except with resistance equal to coil in series.

mercury bridge does not ordinarily break until after contact is made on the opposite side.

The natural frequency of the free armature, wet with mercury, is about 220 cps. The relays have been used at frequencies up to about 350 cps where the drive conditions were individually selected with respect to the phase of impact transients. Fairly well controlled operation without such selection can be obtained up to about 100 cps.

CONTACTS

The relay has been designed with telephone, rather than power applications in mind. The contacts are smaller than those in the previous neutral design, and this appears to be associated with less capability for closure of very high current circuits. The capacity required across the contact to prevent noticeable arcing appears to be about the same as that required in the earlier type:

$$C = \frac{I^2}{10} \text{ microfarads,}$$

but in most cases, larger values than this will be needed to hold peak

voltages to safe values. A small resistance in series with the condenser to limit the closure current is usually necessary. A considerable amount of very satisfactory experience has been had with inductive loads of 0.5 ampere at 50 volts, with 0.5 mf in series with 10 ohms across the contacts.

The contact closure shows no chatter for time intervals of 0.1 microsecond or more.

LIFE

Tests of relays with protected contacts indicate that the only change is a sensitivity change due to wear. Under conditions producing fairly high velocity of contact impact this change is of the order of 5 ampere turns increase in differential ampere turns for a billion operations, the change being roughly proportional to the logarithm of the number of operations.

HORIZONTAL TYPE RELAY

Most of the apparatus in the telephone plant is mounted on vertical panels. The substantially vertical mounting requirement for the relay shown in Fig. 1 tends therefore to be uneconomical from the space standpoint. Fig. 9 shows a modification of the glass enclosed switch which avoids this limitation by being designed to operate with its axis horizontal.

The switch modification consists in adding to the switch a special

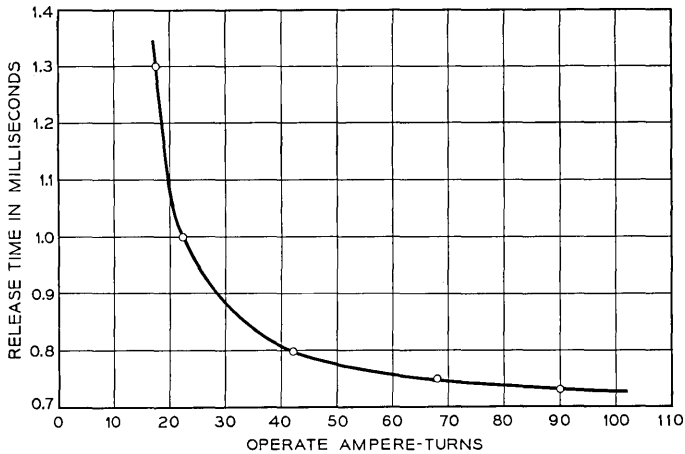


Fig. 8 — Release time versus operate ampere turn adjustment.

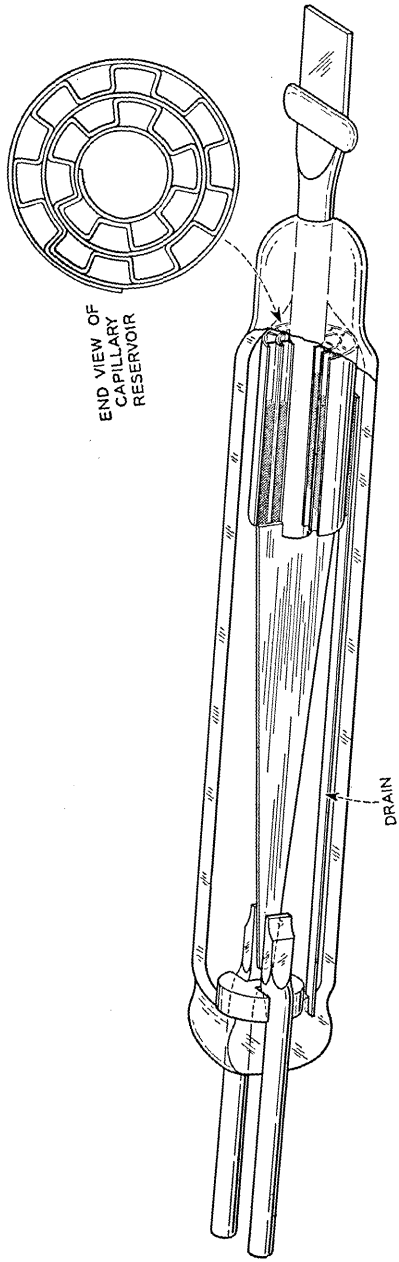


Fig. 9 — Horizontal switch.

capillary reservoir which establishes a negative head in the mercury equal to that obtained by the height of the contacts above the mercury reservoir in the switch shown in Fig. 2. To provide for the return of free mercury to the reservoir, a wet strip of metal is placed along the glass wall underneath the contacts.

The capillary reservoir is essentially a bundle of tubes with walls wet by the mercury. The amount of mercury in the switch is such that the tubes are about half full. The mercury meniscii in these tubes are tangent to the walls and therefore have a spherical surface with a radius equal to that of the tubes. The proper tube radius ($R_1 = R_2 = 0.033$ cm) for

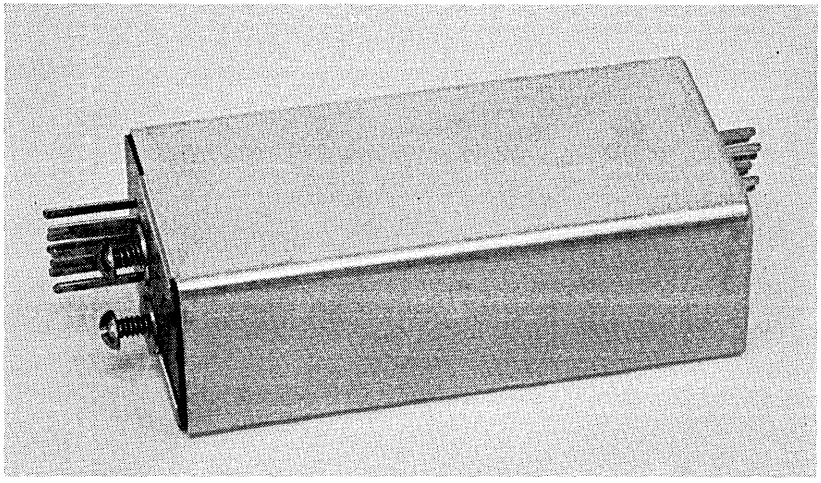


Fig. 10 — Horizontal relay.

the desired negative head ($h = 2.07$ cm) is obtained from equation (3). As the tubes have a uniform bore, the negative head which they establish is not critically dependent on the amount of mercury in them.

The capillary reservoir is made from thin strip, formed and then rolled into a cylinder. It surrounds the metal tube and is welded to the base of the armature. The drain element is sealed into the glass at one end, and, at the other end, is held in contact with the capillary element by the surface tension and negative head of the mercury.

The horizontal switch is an experimental design and has not been put into production. It can be assembled in the housing shown in Fig. 1 for plug connection to a vertical panel. Fig. 10 shows a preliminary model with the new type terminals for wrapped connections.[2] It is adapted for

[2] Solderless Wrapped Connections, B.S.T.J., May, 1953.

mounting on a vertical panel interchangeably with existing polar types. The terminals, which may be up to eight in number, are also brought out at the rear for test purposes.

HIGH-FREQUENCY SWITCHING

The small size, symmetry, and simplicity of the vertical switch structure has been found to be particularly well adapted for incorporation into coaxial structures for high-frequency switching and it is being used for a variety of applications of this kind.

USES

Small scale production of the relays has been started for a few special uses. Designs for larger scale uses are still in a preliminary stage. In general, it is expected that the relay will be well adapted for applications where a transfer contact in combination with low inductance, high speed, high sensitivity, low contact resistance, freedom from chatter, good high frequency switching characteristics, stability, long life, or any combination of these, is required.

APPENDIX

MAGNET STRENGTH VERSUS SENSITIVITY

A simplified representation of the magnetic system of a relay of this general type is shown in Fig. 11. Here the armature is shown working between two magnetic poles N and S, each of which has an area A. The armature position is indicated as a deviation ℓ toward N from the mid position, and is limited in its motion by stops at $\ell = \pm\ell_1$ from moving through the full magnetic gap range, defined by the positions $\ell = \pm\ell_2$.

M_s and M_n are MMF's in ampere turns, introduced by the magnets on either side, with positive values as indicated by the arrows. Similarly, M_c represents the MMF introduced by the operating coil. These values are not those of the magnets and coil in the actual relay. Instead they are assumed to be "Thevenin" equivalent open circuit values of MMF looking away from the working gaps on either side. The gaps are assumed to have the simple geometric dimensions of length $\ell_2 - \ell$ and $\ell_2 + \ell$, and area A. The reluctances looking away from the working gaps are assumed to be represented by the fixed gaps of area A and length $\ell_2 - \ell_1$ on either side.

The magnetic pull on the armature in a gap of this kind is

$$f = \frac{2\pi}{980}(0.1M)^2 \frac{A}{g^2} = \frac{(M)^2}{(124.9)} \frac{A}{g^2} \tag{1}$$

where f is the force in grams g is the gap length and M is the MMF in practical ampere turns across the gap.

In the structure shown, therefore, the net force on the armature from the gaps on both sides is

$$f_m = \frac{A}{(124.9)^2}\left[\frac{(M_n + M_c)^2}{(\ell_2 - \ell)^2} - \frac{(M_s - M_c)^2}{(\ell_2 + \ell)^2}\right], \tag{2}$$

where the values of M are expressed in ampere turns.

If $M_n = M_s$ this can be converted to

$$\frac{f_m}{f_{ml}} = \left[\frac{\left(1 + \frac{M_c}{M_n}\right)^2}{\left(1 - \frac{\ell}{\ell_2}\right)^2} - \frac{\left(1 - \frac{M_c}{M_n}\right)^2}{\left(1 + \frac{\ell}{\ell_2}\right)^2}\right], \tag{3}$$

where $f_{ml} = \left(\frac{M_n}{124.9}\right)^2 \frac{A}{\ell_2^2}$.

Fig. 12 shows curves of $\frac{f_m}{f_{ml}}$ versus $\frac{\ell}{\ell_2}$ for various values of $\frac{M_c}{M_n}$. Only one quadrant is shown as the system is symmetrical. It provides a convenient means for analysis of this type of relay.

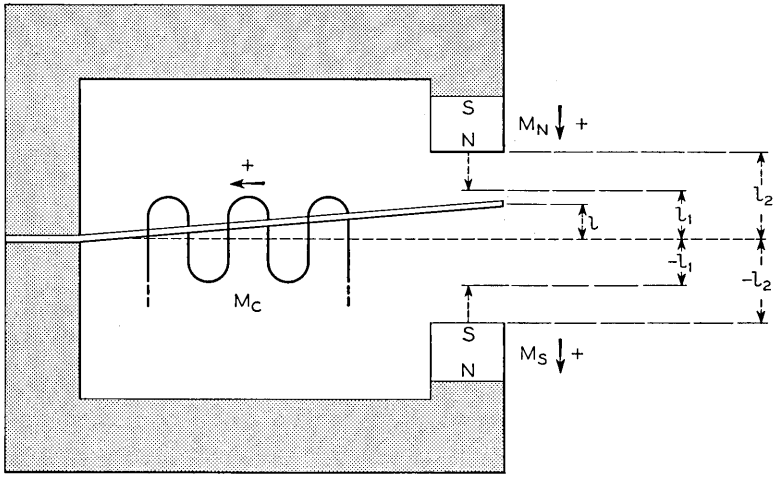


Fig. 11 — Simplified representation of relay.

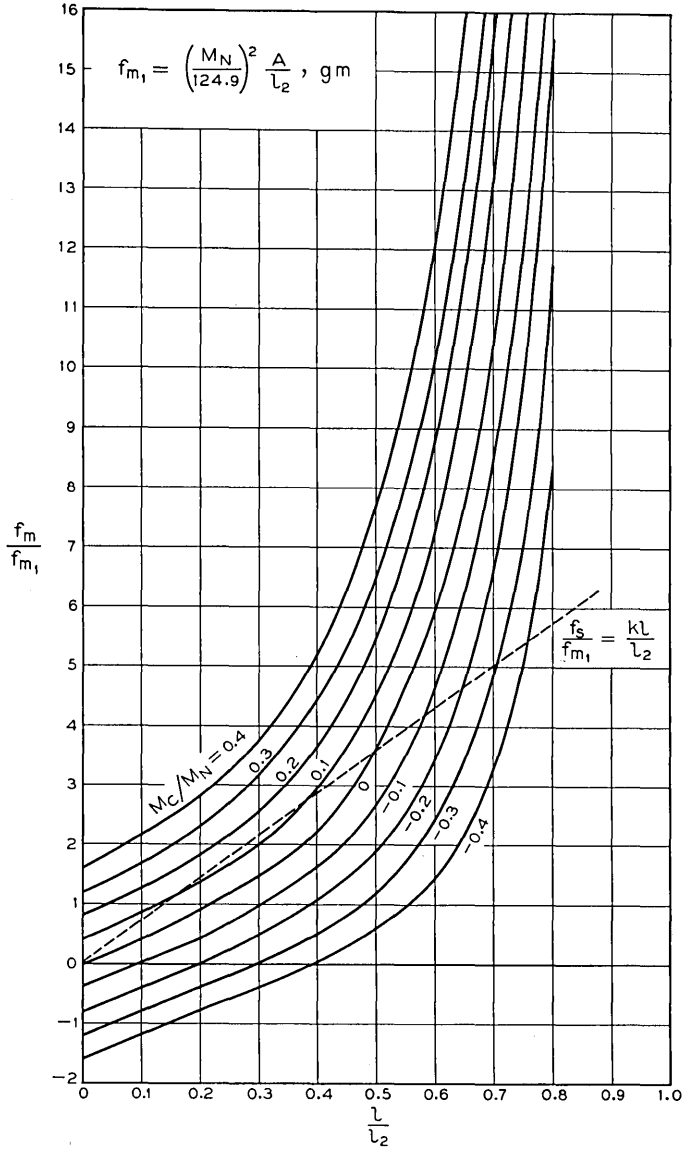


Fig. 12 — Balanced gap relay chart.

Also shown on the figure is a typical spring characteristic

$$\frac{f_s}{f_{ml}} = \frac{K\ell}{\ell_2}. \tag{4}$$

If the stops ℓ_1 and ℓ_2 are placed at points of intersection of the spring characteristic with the magnetic characteristic for $M = 0$, (in this case at $\ell_s/\ell_2 = 0.48$), we have a condition that corresponds to a hypothetical relay of infinite sensitivity. That is, the armature can be released from contact with either side with an infinitesmal coil input, and, if there are no magnetic or mechanical losses in its travel, it will just swing to the opposite side without any change in the total energy of the system. Let us define the magnet strength for this condition as M_0 and let us assume that $M_c = 0$ for operation in either direction with this magnet strength.

Assume now that the magnet strengths on each side are increased equally to the values

$$M_{n1} = M_{s1} = M_0 + \Delta. \tag{5}$$

For values of ℓ_1/ℓ_2 near 1 this change can be balanced out by a coil input of about the same amount, as practically all of the pull on the armature would be from the nearer pole. For values of ℓ_1/ℓ_2 near 0 the effect of a coil input will be equal and aiding in both gaps. The coil inputs required to just operate from the N and S poles, defined as M_{cn1} and M_{cs1}, respectively for this particular case, will be

$$M_{cs1} = -M_{cn1} = p\Delta = p(M_{n1} - M_0) = p(M_{s1} - M_0), \tag{6}$$

where p is a value between 1 and 0.5. Values for individual cases can be worked out with reference to the curves for various values of M_c/M_n in Fig. 12. For the case shown, where $\ell_1/\ell_2 = 0.48$, p is about 0.8.

An adjustment in accordance with (6) is thus a balanced one with a spread of $2p\Delta$ between the two sensitivity values, the amount of spread increasing with increase in the strength of the two equal magnets M_{n1} and M_{s1} above the value M_0.

A general type of adjustment, including all possible combinations of the two sensitivity values, can be obtained by adding a suitable value B to a balanced pair of sensitivity values in accordance with equation (6). These general sensitivity values would then be

$$\begin{aligned} M_{cs} &= M_{cs1} + B, \\ M_{cn} &= M_{cn1} + \mathrm{B}. \end{aligned} \tag{7}$$

This is the type of change that would be produced by a bias of $-B$

ampere turns in an auxiliary coil. The equivalent of such a bias would be obtained by decreasing the strength of the N magnet and increasing the strength of the S magnet, each by the value B. The general magnet strength values would then be

$$\begin{aligned} M_n &= M_{n1} - B, \\ M_s &= M_{s1} + B. \end{aligned} \tag{8}$$

Combining equations (5), (6), (7) and (8) to eliminate M_{cs1}, M_{cn1}, M_{n1}, M_{s1} and B, we obtain the general relations between sensitivity and magnet strength as

$$M_{cs} = p\left(\frac{M_n + M_s}{2} - M_0\right) + \frac{M_s - M_n}{2}, \tag{9}$$

$$M_{cn} = -p\left(\frac{M_n + M_s}{2} - M_0\right) + \frac{M_s - M_n}{2}. \tag{10}$$

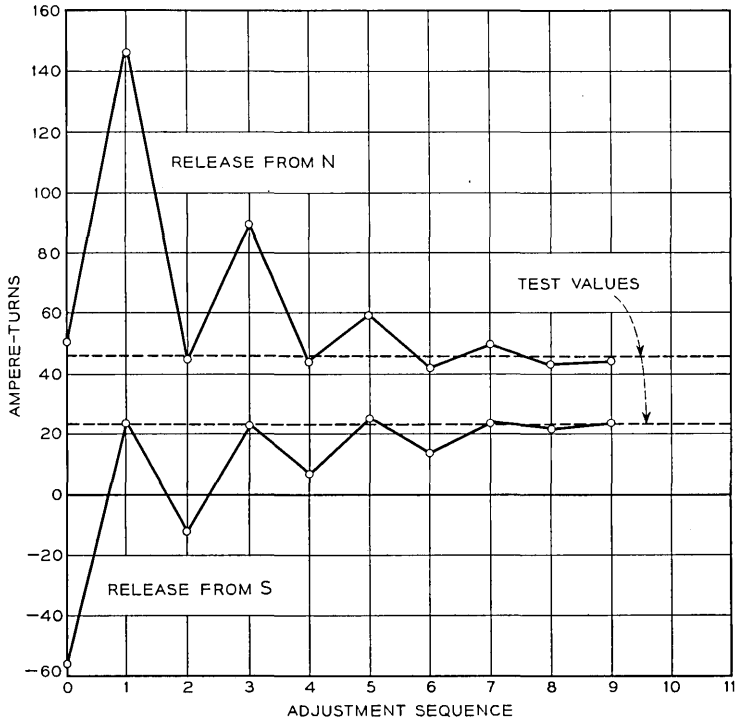


Fig.13 — Sequence of sensitivity values obtained in adjustment.

ADJUSTMENT OF SENSITIVITY

The procedure used to adjust the relay to any desired pair of sensitivity values is as follows:

1. Magnetize the two magnets fully by a flux directly across them.
2. With the armature starting from one side, progressively decrease the magnetization on that side in small steps until the armature just operates to the other side on the final current value desired for this direction of operation.
3. With the armature starting on the other side, progressively decrease the magnetization on that side in small steps until the armature just operates from that side on the final current value desired for this second direction of operation.
4. Repeat 2 and 3 alternately until the relay just operates in both directions on the two current values desired.

Within limitations such as the size of demagnetizing increments used per step, this procedure results in an adjustment to the two test values used. Let us consider why this result is obtained.

If we should substitute $p = 1$ in equations 9 and 10 we would obtain

$$M_{cs} = M_s - M_0 ,$$
$$M_{cn} = M_1 - M_n .$$

That is, in such a case, the two adjustments would be independent of each other and the above process would require only one adjustment of each magnet. In the more general case, where the force on the armature is affected by both poles, each adjustment on one side results in a greater magnet strength at the pole being adjusted than is required for the final adjustment, because of the pull from the opposite pole, the amount of which is greater than normal because that pole has not been reduced to its final magnet strength. Thus the final adjustment is normally reached through a number of alternations between the two sides, which progressively approach the final pair of sensitivity values.

The sequence of sensitivity values obtained in a typical adjustment of this kind is shown in Fig. 13.

Dynamic Measurements on Electromagnetic Devices

By M. A. LOGAN

(Manuscript received July 6, 1953)

A sampling switch with adjustable make and fixed break times can be used to obtain dynamic measurements of reciprocating phenomena. A test set has been developed using this principle to measure the flux-time, current-time, displacement-time, and velocity-time response of electromagnets and similar devices. The tested device is steadily cycled. A dc instrument is switched in by the synchronous control at any preselected instant in the cycle, and out when a steady reference condition has been reached. The steady reading of the instrument is proportional to the value, at the closure instant, of the variable being measured. The instrument switching is controlled by an electronic timing system. This system operates mercury contact relays, which do the actual switching. For the displacement-time and velocity-time measurements, an optical transducer with associated dc amplifiers is added. The design of these devices is described. The results of an investigation of dynamic flux rise and decay in solid core electromagnets are reported. Modified first approximation equations are developed to give a better representation of eddy current effects.

Introduction

The application of common control methods to telephone switching systems has led to the widespread use of high-speed relays. The actuation time of these relays is affected by many parameters such as the power supplied, how far the armature has to move, the mechanical work the armature has to perform during its motion, the winding design, the magnetic structure, and eddy currents introduced in the magnetic members caused by the application of current to the winding. The eddy currents act to oppose a magnetic flux change and hence retard a building or decay of flux. This causes the actuation time to be increased compared to a relay without such effects. An analytical determination of the development and effects of eddy currents can be made for simple sym-

metrical magnetic structures having an infinitely long or torus shaped round or rectangular cross-section, assuming linear magnetization characteristics. However, for relay-like structures having air gaps, leakage flux which only partly completes its circuit through the magnetic material, varying cross-section so that boundary conditions become complicated, and non-linear magnetic properties, an analytical approach becomes unmanageable.

For a fundamental study and direct measurement of eddy current effects, a test set has been developed to measure the dynamic flux rise and decay characteristics of relays, and similar structures. This test set is electronically operated on a synchronous switching principle. It displays on an ordinary dc instrument, as a steady reading, the instantaneous flux obtaining at any selected time after energization or de-energization of the magnet.

The application of the test set is restricted to devices which operate under conditions of suddenly applied or removed dc voltage and which can be cycled between these two conditions until the dc instrument reaches its steady state reading.

This article will present the theory of the basic synchronous switching circuit and the relation applying between the dc instrument reading and the instantaneous flux linkages, a description of the electronic control circuits, and measurements of dynamic flux rise and decay using solid core electromagnets.

Supplementary additions to the basic circuit are available using the same principles, to measure current-time, displacement-time, and velocity-time curves of reciprocating devices. The first fundamental measuring set using the switching principle, was built by E. L. Norton[1] in 1938. The earlier set used a synchronous, motor driven, phase adjustable commutator, to perform the switching. A limitation of the earlier set in how fast it could be operated, combined with brush wear and chatter troubles led to the development of the new electronically controlled set employing sealed mercury contact relays to perform the switching.

Part I — Description of Fluxmeter System

BASIC MEASURING CIRCUIT

A schematic of the basic measuring circuit is shown in Fig. 1. A battery switching contact indicated as A, applies and removes voltage to the magnetizing winding of the electromagnetic device under test with a 50 per cent duty cycle. The time of one complete cycle is indicated as T. The electromagnetic device is equipped with a search coil of N turns, having dc resistance, including wiring, of R_c ohms. A B contact, synchronously switched at instants to be described later, connects a dc microammeter to the search coil. A damping contact C connects the instrument to a resistance, preset to the same value as the search coil, when the instrument is not connected to the latter, thus providing the same instrument damping at all times.

The A timing cycle is shown schematically for one interval T. Next below is shown the cycle for the B contact, followed by that for the C contact. The B contact always opens and the C contact closes just before the A contact closes. The point of closure of the B contact, indicated as t_1 may be adjusted to occur at any time during the cycle T. The cycle time T is chosen sufficiently long so that the flux in the electromagnet has sufficient time to reach a substantially steady state during both the closed and the open intervals.

The flux rise and decay have the general characteristics shown as the φ curve, while below it is the induced voltage in the search coil, exactly given by Lenz's Law

$$e = N\frac{d\varphi}{dt}. \tag{1}$$

The first property of this last curve to observe is that the positive and negative areas are exactly equal. Otherwise a dc instrument connected permanently to the search coil would show a dc current flow. The second property is that the area of the curve up to some time t_1 is proportional to the instantaneous flux at the time t_1. Finally, this unshaded area is numerically equal but opposite in sign to the shaded remainder of the area, taking into account the signs of the two components of the remainder. The operation of the test set depends upon measuring this shaded area.

Let the flux linking the search coil of N turns at any time be φ. Assume that the microammeter has a resistance R_m and inductance L_m. The current flowing in the meter mesh while the B contact is closed is

$$N\frac{d\varphi}{dt} + L_m\frac{di}{dt} + (R_m + R_c)i = 0. \tag{2}$$

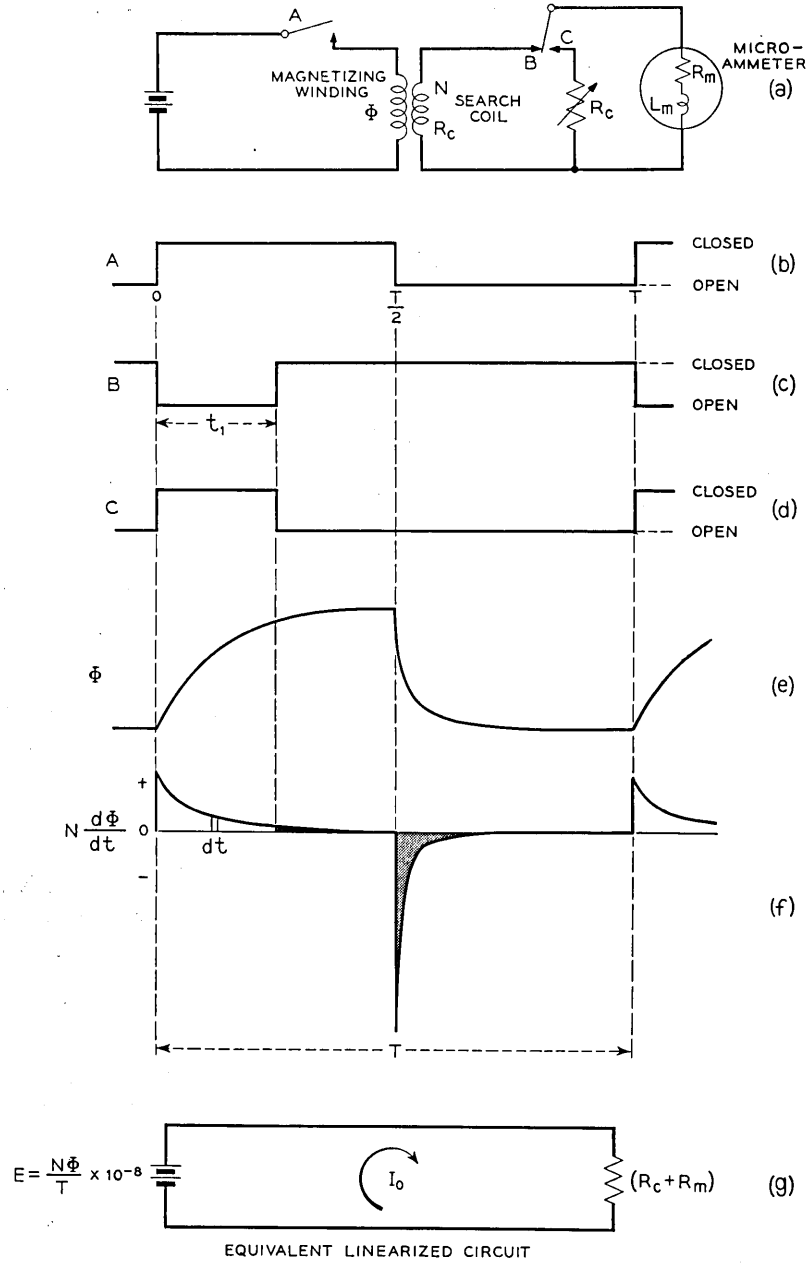


Fig. 1 — Fundamental operation of dynamic measuring set.

The dc instrument will read the average value of current for the cycle. As the period of the cycle is T, and t_1 is the chosen delay for closure of the B contact measured from the time of the A closure, the direct current indicated by the instrument will be:

$$I_0 = \frac{1}{T}\int_{t_1}^{T} i\,dt = -\frac{1}{(R_m + R_c)T}\int_{t_1}^{T} \frac{d}{dt}(N\varphi + L_m i)\,dt$$
$$= \frac{(N\varphi + L_m i)_{t_1} - (N\varphi + L_m i)_T}{(R_m + R_c)T}. \tag{3}$$

Now the product $L_m i$ at t_1 when the B contact is closed must be zero. Furthermore, if the period T is long in comparison with the time of rise and decay of the flux, and with the time constant L/R of the measuring circuit, the product $L_m i$ is also negligible at $t = T$. We then have:

$$\Phi_{t_1} - \Phi_T = \frac{RTI_0}{N} \times 10^8 \text{ Maxwells}, \tag{4}$$

where R is now the total resistance of the measuring circuit. The factor 10^8 has been added to convert from practical units to c.g.s. units. The flux Φ_T is the residual flux, so that if this is taken as a reference value, the value of the flux at the time of closing the B contact is simply

$$\Phi = \frac{RTI_0}{N} \times 10^8 \text{ Maxwells}. \tag{5}$$

It may be objected that the flux in equation (2) is made up not only of the flux to be measured but also of a component due to the current in the measuring circuit. This is true, and the flux linking the search coil changes differently during the time the B contact is closed compared to that without the instrument circuit. Note, however, that we are not concerned with how the flux varies between the times t_1 and T but only with its value at the limits. Since the flux at t_1 is that which has been established with the search coil circuit open, the one requirement is that the constants be such that the current in the measuring circuit be substantially zero just before T. This can be verified through the measurements themselves by noting whether there is any change in measured flux in this interval. If there is, then the cycle time T can be increased until this requirement is met.

It will be convenient later to regard the switched meter as a linearized circuit composed of an equivalent dc voltage

$$E = \frac{N\Phi}{T} \times 10^{-8} \text{ volts} \tag{6}$$

in series with the resistor R, causing a dc current I_0 to flow, Fig. 1 (g).

MEASUREMENT OF CURRENT

The instantaneous value of current in a circuit being cycled by the A contact is measured by inserting the primary of a toroidal air core transformer in series with the circuit as shown in Fig. 2. The secondary is connected to the meter through the B contact. The best design consists of a secondary occupying half the winding volume and wound to the same resistance as the meter. The primary can be wound in two sections, half inside and half outside of the secondary to provide maximum mutual inductance, of a wire coarse enough to provide the number of turns for a convenient meter deflection. In an air core transformer, the magnetic circuit is linear and the flux exactly proportional to the current, a situation that does not exist when magnetic materials are present. Because of this linearity, the analysis can be made in terms of inductances and currents rather than flux linkages and currents.

By definition, the voltage induced in the secondary of a linear transformer is related to the primary current, through the mutual inductance M by the relation:

$$e = M \frac{di}{dt}, \tag{7}$$

which corresponds exactly to (1) with M substituted for N, and i for φ. Hence, the instantaneous current at time t_1 is given by the relation

$$i = \frac{RTI_0}{M}, \tag{8}$$

where the other symbols have the same significance as before. This relationship makes the calibration of the mutual inductance easy by setting up a battery and resistor for the primary circuit and measuring the steady state current with an ammeter. Then when the A contact is being switched and the instrument current I_0 is read, corresponding to the known final primary current i, relation (8) can be solved for M.

MEASUREMENT OF FLUX USING A BRIDGE CIRCUIT

The search coil method of measurement just described is preferred because of the absence of any dc voltage in the meter circuit. However,

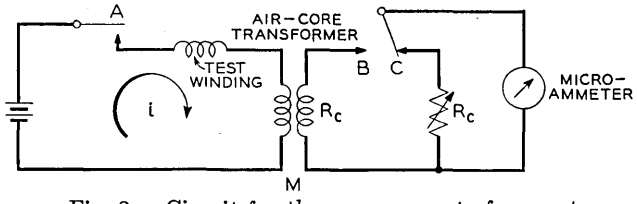


Fig. 2 — Circuit for the measurement of current.

sometimes the use of a search coil is inconvenient, or for other reasons it is desired to measure the flux linkages in the magnetizing winding itself. The test set includes a bridge circuit shown in Fig. 3 for such measurements. With the contacts A and B closed and the bridge in the flux rise arrangement, the bridge is first balanced for no current in the instrument, by adjustment of R_d. The same instrument is used for setting the balance as later used for the flux measurement. The damping resistance is next set to the value facing the instrument, namely,

$$R_c = \frac{K}{K+1}(R_0 + R_d). \tag{9}$$

If then the contacts are operated with the period T, the instrument will read a current I_0, and the flux-turns at time t_1 during rise will be

$$N\Phi \times 10^{-8} = R_m I_0 T\left[1 + K\left(1 + \frac{R_0}{R_m} + \frac{R_d}{R_m}\right)\right] - L_1 i, \tag{10}$$

where N is the number of turns of the winding, Φ is the average flux per turn, and the other quantities are indicated on the drawing. The term $L_1 i$ is a correction term, usually negligible, involving the self inductance of the primary of the air core current transformer and i the instantaneous

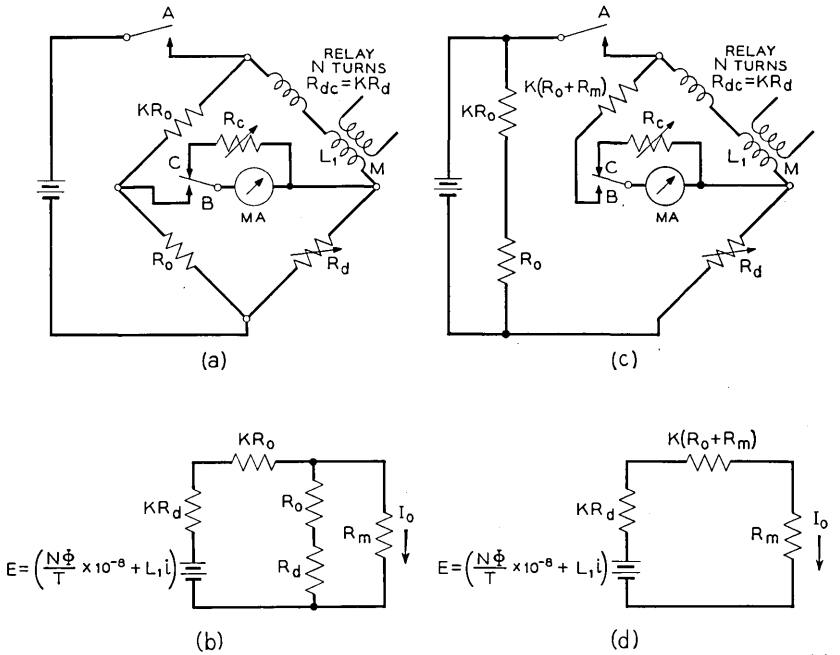
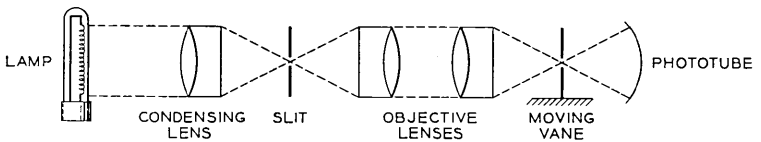


Fig. 3 — Bridge circuit for the measurement of flux — (a) and (b), rise; (c) and (d), decay.



OPTICAL SYSTEM

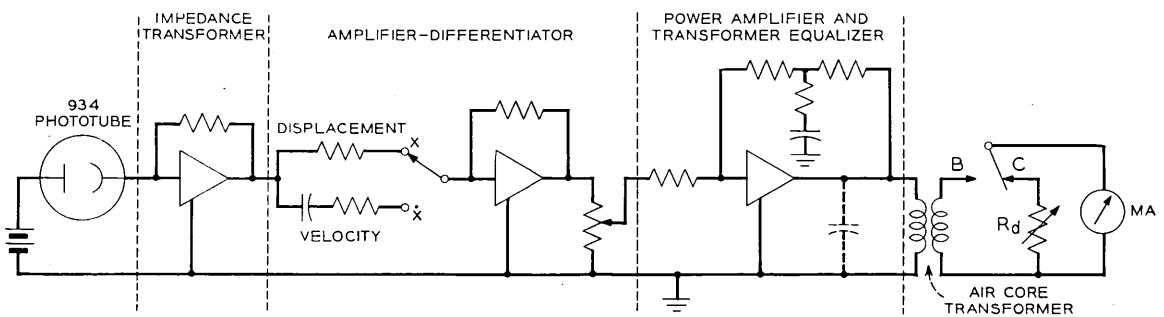


Fig. 4 — Block diagram of photocell amplifier system.

current at the time t_1; i is determined by a separate measurement described above. If its value is not needed, the air core coil can be omitted from the circuit.

For flux decay measurements made after the A contact has opened, the bridge circuit has to be modified to remove the resistance R_0 and KR_0 from being across the electromagnet and affecting its flux decay. They are transferred to the battery side of the A contact as shown in Fig. 3(c), to maintain the same battery drain. This avoids any error due to the internal resistance of the battery, which would otherwise cause the final flux at the end of a rise test to differ from the initial flux at the beginning of a decay test. The added resistor $K(R_0 + R_m)$ is connected in series with the meter to make the expression for the flux-turn linkages the same as before. For decay measurements the damping resistor is set at the value

$$R_c = K(R_0 + R_m + R_d). \tag{11}$$

In Figs. 3(b) and 3(d) are shown the linearized equivalent circuits for rise and decay respectively, from which equation (10) above can be derived conveniently.

MEASUREMENT OF DISPLACEMENT AND VELOCITY

An optical probe is provided, in which the amount of light falling on a photocell is controlled by the relative position of two flags, one cemented on each of the relatively moving parts to be studied, such as an electromagnet. The change in output current from the photocell thus is proportional to the displacement of the armature with respect to the core, one flag being on each. A block diagram of the system is shown in Fig. 4. Amplifiers effective from dc up to a frequency determined by the resolution required, with substantially no phase shift, deliver a current into the primary of an air core transformer proportional to the instantaneous displacement of the armature. By virtue of the linearity of this transformer, the flux developed is proportional to the displacement. Thus by operating the electromagnet with the A contact and connecting the secondary of the transformer to the B contact, the instrument gives a dc reading proportional to the armature displacement at the time the B contact closes. The total displacement can be measured statically. The instrument reading at a time after complete operation corresponds to this known displacement, permitting the scale to be converted to a displacement scale.

By changing one amplifier input resistor to a capacitor, the amplifier

becomes a differentiator, in which the output current is proportional to rate of change of input voltage. With this one change, the instrument reads the instantaneous velocity of the armature. Under these conditions, the second amplifier differentiates the input voltage once, the output transformer differentiates a second time, and the dc instrument integrates once, leaving a net result of one differentiation.

Description of System Components

FLUXMETER

General

In the following descriptions of the several components of the system, the general characteristics required are considered, the specific deviations from ideal are determined, and an evaluation of the measurement errors is made. The description starts with the dc instrument followed by the associated vacuum tube filter. Then the heart of the system, the contact switching circuit itself is considered. Following this is the timing impulse generating circuit, the counting rings, the time selector, coincidence circuits, memory, and relay control circuits. These elements make up the fluxmeter proper.

The auxiliary circuits for displacement-time and velocity-time measurements are the concern of the next part of the paper. The components are the optical system, the photocell amplifier, the amplifier-differentiator and finally the output amplifier.

The last section first shows typical measurements on a telephone type relay. Then a description of a more fundamental study of dynamic flux rise and decay in solid core electromagnets is given. This study has led to two new first approximation equations for dynamic flux rise and decay, as will be seen.

Dc Instrument and Effect of Damping Resistance

The dc instrument used for a majority of the measurements is a heavily damped 50.0-ohm, 200-microampere full scale instrument, with a ½ per cent of full scale accuracy. The open circuit decay time constant is about 2 seconds, and with a shorted winding about 4 seconds. On Fig. 5 is shown a plot of the decay time constant referred to a short circuit, versus the damping resistance referred to the meter resistance.

An evaluation of the small error introduced by not providing the damping resistor is as follows. Consider a square wave flux pattern produced by an ideal electromagnet, with zero winding time constant or

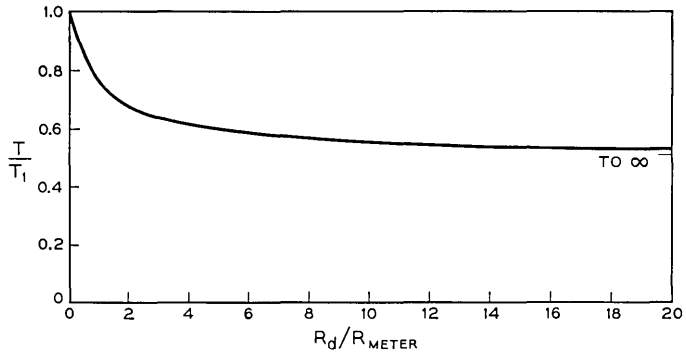


Fig. 5 — Decay time constant of DC instrument.

eddy currents. The induced voltage in the search coil then consists of two equal and opposite impulses occurring respectively at 0 and T/2. To measure the constant flux, the switching epoch can be anywhere between these two values. The switching omits the impulse at zero time. For limiting cases t_1 can approach either 0 or T/2. The instrument reading of course should be independent of the choice, as the flux remains at the constant maximum throughout this interval. The instrument receives only the decay impulse, and thus there will be an average dc component.

For the case of t_1 approaching zero, the instrument is always connected to the search coil, which usually is of negligible resistance and the short circuit meter decay time constant applies. Just before a succeeding impulse, the meter will have decayed to a relative value of:

$$e^{-T/T_1} \approx 1 - \frac{T}{T_1}, \tag{12}$$

where T is the cycle time as before and T_1 is the short circuit decay time constant. The pointer next abruptly rises to its original maximum value, followed again by another equal decay in a cyclic manner.

For the case of t_1 approaching T/2, the instrument is connected to the the search coil only half the time, being open circuited for the remainder of the time, if no damping resistor is provided. It thus decays at two different rates following the impulse. The first half is the same as before but the second half is under the condition of open circuit. For small errors the relative decay will be:

$$e^{-T/2T_1}e^{-T/2T_2} \approx 1 - \frac{T}{2}\left[\frac{1}{T_1} + \frac{1}{T_2}\right], \tag{13}$$

where T_2 is the open circuit decay time constant.

The search coil impulse is the same in each case, and for small errors, the instrument will rise the same amount for either condition followed by equal decays. The two maximum values M_1 and M_2 are not equal but are related by:

$$M_1 \frac{T}{T_1} \approx M_2 \frac{T}{2}\left(\frac{1}{T_1} + \frac{1}{T_2}\right) \tag{14}$$

$$\frac{M_1}{M_2} \approx \frac{1}{2}\left(1 + \frac{T_1}{T_2}\right) \tag{15}$$

The ratio of the two instrument readings thus is a function of the ratio of the two instrument decay time constants. For the case of $T_1 = 4$ sec., $T_2 = 2$ sec., the limiting error is 33 per cent. This shows that without a damping resistor, the error can be many times the ½ per cent instrument error, but that the timing for switching the meter damping is not critical. That is, small time gaps with the damper off will introduce an insignificant error. In a sense, the omission of the damper is equivalent to using an electronic integrator with finite internal gain.

Vacuum Tube Filter Circuit[1]

The above discussion brings out the fact that visible motion of th instrument pointer without a filter occurs, increasing as the cycle time ie lengthened. From equation (12) and using $T = 0.1$ sec., and $T_1 = 4$ sec.s the amplitude would be about 2½ per cent of full scale. Reading the center of the vibration is easy for cycle times T less than 0.1 second and for such measurements the filter is not used. Many measurements, however, require a slow pulse rate and even a sluggish instrument will follow the pulses to such an extent that an accurate reading is impossible.

Fig. 6 is a schematic of a vacuum tube circuit which serves as a very efficient low pass filter. This filter must fulfill a variety of very stringent requirements, chief of which may be listed:

(a) Low loss to direct current.

(b) High loss, probably exceeding 30 db to frequencies above one cycle per second.

(c) A constant resistance input.

(d) Ability to suppress peaks of relatively high voltage of the order of several hundred or thousand times the dc voltage being measured.

(e) No effect other than one which may be accurately calculated on the dc reading.

(f) The circuit should not greatly increase the time required for the instrument to reach a steady reading.

Requirement (c) may call for some comment. It was pointed out earlier that the method requires the time constant of the measuring circuit to be small in comparison with the period of the pulses. A filter built in the ordinary way to have high suppression to pulses of period T would not have a time constant small in comparison with this figure. The require-

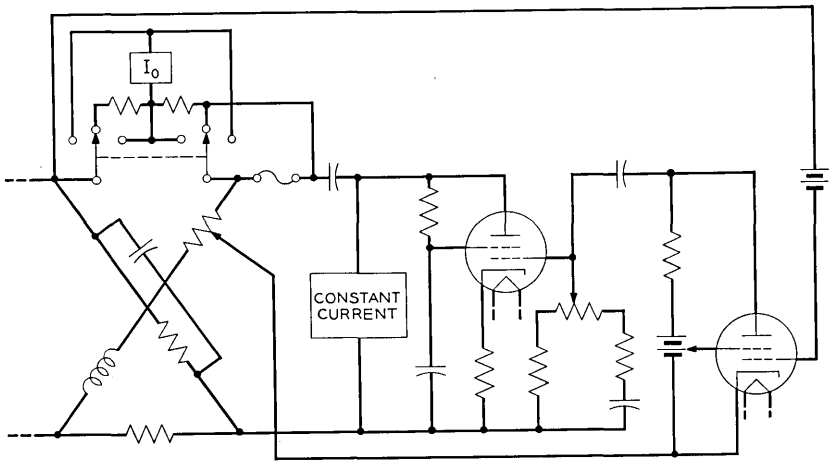


Fig. 6 — Vacuum tube low pass filter circuit.

ment of constant resistance input is a convenient way of expressing the necessity of fulfilling this condition.

The original work on the system used was done by R. F. Wick, and the features to be described are due to him and E. L. Norton. From left to right in Fig. 6, the elements are as follows: a balanced impedance bridge containing the ammeter in one arm, a blocking condenser, a constant current high impedance power supply for the power tube, and a two stage amplifier with an interstage phase adjusting circuit. The three-point double-pole key when in the normal position removes the meter (50 ohms) and substitutes a 50-ohm resistor. When off-normal, the meter is connected in either polarity.

The operation of the circuit is best understood by assuming the reactance elements to be omitted from the bridge and the contact on the potentiometer forming one diagonal to be at the lower left. The input to the amplifier is then directly across the line and any feedback is elimi-

nated, since the bridge is balanced. Any alternating component applied will be amplified and sent back through the meter in the opposite polarity to that coming directly from the line. If then the gain in the amplifier is correct and its phase shift nearly zero, the alternating component through the meter will be greatly reduced.

The bridge provides the constant resistance feature at the input, since the output of the amplifier can have no effect on the input impedance. It does however necessitate a loss from both line and amplifier to the measuring instrument.

Practically, the difficulty in design is in securing an amplifier with enough peak output capacity and negligible phase shift at frequencies from one cycle up. This has been accomplished by several methods. Most of the phase shift is introduced by the plate condenser in the last stage. The effect of this is reduced by the constant current power supply.

The gain of the amplifier is controlled by feedback secured by the potentiometer in the diagonal arm of the bridge. The gain is sufficient so that a substantial amount of feedback may be used with a consequent further reduction in phase shift.

The final phase compensation is secured by the interstage potentiometer. The effect of this is illustrated in Fig. 7. The lower curve is the net phase shift of the amplifier without the interstage circuit. The upper curve is the phase shift which may be introduced at the grid of the second stage. By proper adjustment, these may be made to compensate each other down to a frequency of one or two cycles. The circuit constants are such that the final adjustment of low frequency phase may be made with a negligible effect on high-frequency gain. It may be of interest that if the bridge is unbalanced by shorting the lower 50-ohm resistance, an error of 20 per cent or more is introduced in the dc reading due to the extremely large effective reactance of the feedback amplifier. If the bridge is unbalanced by shorting the instrument the amplifier is quite likely to "motor-boat" and blow the fuse used for protection. It is for this reason that the instrument key may be set to replace the instrument by a 50-ohm resistor. When in this position the fuse is also shorted to avoid needlessly blowing it due to the high surges which frequently occur when the amplifier is turned on and the condensers start to charge.

The filter as described, omitting the reactance elements from the bridge, would be entirely satisfactory within the power capacity of the amplifier. With certain measurements, notably those of velocity, the peaks of the wave as applied to the filter, which in this case would be proportional to acceleration, are so high that no reasonable amplifier would be able to handle them. These knifelike peaks however are so sharp that they may

easily be removed by reactance elements added to the bridge as shown. It will be noted that the bridge is still of constant resistance, and is still balanced at all frequencies.

Three adjustments are provided on the amplifier: one controls the feedback, a second the phase correction, and the third the plate current. The amount of suppression is controlled by a three position key which may be used to cut down the sizes of the capacitors in the circuit in a ratio of two and four to one. In cases where one of the higher pulse rates is used, adequate suppression and somewhat faster readings may be secured by using smaller capacitors. With the maximum suppression, two cycles may be reduced to such an extent that it has no detectable

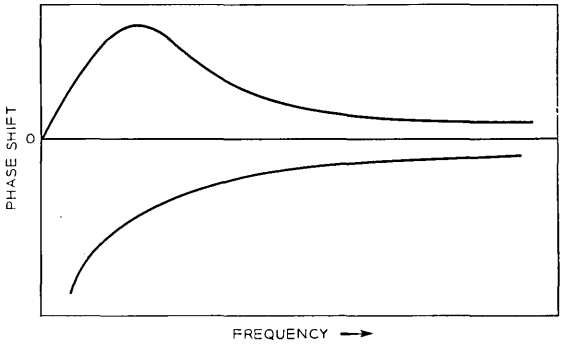


Fig. 7 — Illustration of method for phase compensation of amplifier.

effect on the meter pointer. Accurate measurements at one cycle may be made, although there is a slight motion of the pointer.

Analytical studies have indicated that the transient response of the circuit depends to a large extent on the ratio of the interstage coupling capacitor to the output capacitor and moreover that there is an optimum ratio of the phase compensating capacitor to the other two. In altering the amount of suppression therefore, the ratios of the three capacitors are held constant.

The Switching Circuit

Requirements. The basic feature of this measuring system is the switching circuit, consisting of the A, B and C contacts. The requirements for these contacts are: (a) Negligible dc resistance, (b) No contact chatter, (c) Contact potential less than 10 microvolts, and (d) Stability of operation of 20 microseconds. Consideration of these facters led to the selection of mercury contact relays as the basic switching elements. The choice

of relays leads to additional requirements: (e) Substantially equal operate and release times, (f) A make and a break contact, and (g) Substantially zero transfer or bridging time when operated or released.

The first requirement is to avoid correction terms in the test set equations. The second is to make the test set operation conform to the differential equation applying. The third depends upon the fact that meters having dc resistances of the order of 50 ohms are used, and no error deflection due to contact potential effects should be present. The stability requirement results from an objective of studying flux phenomena in intervals as small as 25 microseconds. The added relay requirements will be developed during the switching circuit description. All except the last can be met by individual relays. The last one has not been met by individual relays, but by using contacts of two relays actuated simultaneously, and adjusting their relative timing by winding shunts, the contact switching interval can be reduced to a few microseconds.

Contact Operation. The switching functions required were shown in Figs. 1 (b), (c) and (d) where the interval t_1 can be as low as 25 microseconds or almost as long as T. The A switch in unaffected in duty cycle, but the B and C switches vary from 0 to 100 per cent duty cycle. The operate and release times of relays are affected by their duty cycle and for low or high values become very erratic. To avoid this effect all relays are operated with a 50 per cent duty cycle, and the switching is accomplished by combinations of relay contacts. The epoch of the B relay cycle can be adjusted by manual selection, anywhere between 0 and $T/2$. This is shown schematically in Fig. 8 for the make contacts. The break contacts of course show exactly the reverse behavior. The C make contact is fixed in phase and is set to open just before the A contact closes.

To produce a B cycle for flux rise shown in Fig. 1 (c) where t_1 is less than $T/2$, a parallel connection of the B and C make contacts is used. To pro-

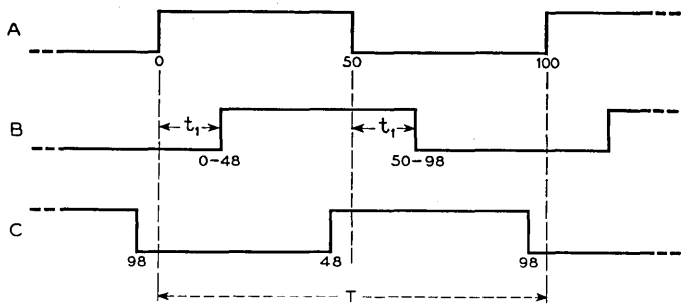


Fig. 8 — The operation of the switching make contacts.

duce a B cycle for flux decay where t_1 is greater than $T/2$, a break contact on the B relay is used in series with the C make contact. Note that in all conditions the meter circuit closure is performed by a B contact and the meter circuit open is performed by the C contact, in fixed phase relationship to the A contact. The addition of the switched meter damping resistor is accomplished by other contacts on the same relays.

A minimum contact circuit[2] to perform all these switchings is shown in Fig. 9 but was found to be unsatisfactory because of failure to meet requirement (g) above. Only a single transfer switch is necessary to change from flux rise to decay, by transferring the wire marked "x" from the search coil to the damping resistance.

The relays are shown in the released position with the break contacts closed. For the flux rise circuit, contact B can have a momentary open interval when it operates. However, it must not bridge or the damping resistor will momentarily be across the instrument at the initial connection, causing an error. But after contact C operates and then B releases at $(T/2 + t_1)$, an open interval at B will momentarily disconnect the instrument. This would cut out part of the current which must be integrated, causing an error. Thus there are conflicting requirements for flux rise measurements, and transfer contacts of a single relay cannot meet both simultaneously. For flux decay, the B contact cannot bridge because when it releases to connect the instrument, a bridging would momentarily connect the damping resistor across the instrument at the instant of greatest interest, causing an error. The requirements for the C contact are of the same type but less stringent, as the flux is never changing rapidly whenever it is switched.

The foregoing discussion was directed toward establishing the basis for the requirements (f) and (g) above. The former is because of the need for phase reversals in contact operation to perform the necessary

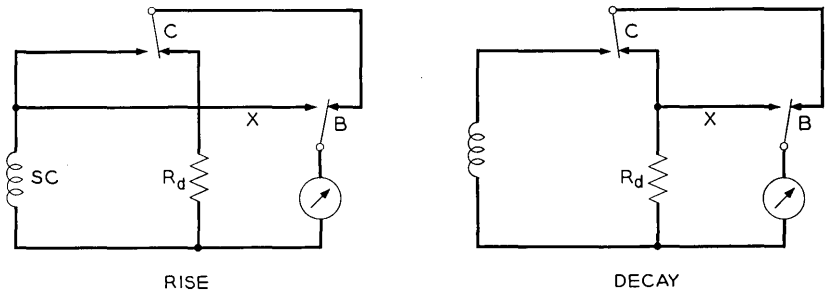


Fig. 9 — Unsatisfactory minimum contact circuit because of transit time requirements.

instrument and damping resistor switching. The latter is to eliminate errors due to momentary circuit configurations not in accordance with the basic equation (3).

The final circuit developed is shown in Fig. 10, which uses two B relays, called $B1$ and $B2$, a single C relay, and compatible timing requirements for the two B relays whose windings are connected in series. In all cases $B1$ shall be faster than $B2$. The former controls the damping resistor while the latter switches the instrument. A composite circuit is drawn, with the cross marks indicating a closed contact for the function being measured, open otherwise. The change from rise to decay in this circuit requires four switch contacts compared to two for the minimum contact circuit. As before, the relay contacts are all shown in the unoperated position. A preliminary sorting of the $B1$ and $B2$ relays is required and is based on operate and release time measurements. The faster relay is assigned as $B1$, the other as $B2$. $B2$ is further slowed down selectively to match the A relay as to operate and release time by the shunts and diode shown in Fig. 10.

$B2$ can be lined up with the A contact both for simultaneous closure of the make contacts when $t_1 = 0$ and for simultaneous closure of the

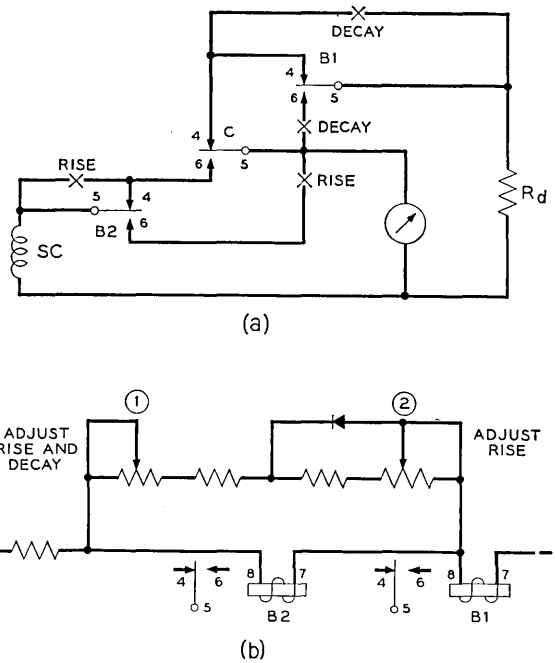


Fig. 10 — Final schematic of basic contact circuit.

break contact of $B1$ with the opening of the make contact of A. The A relay winding has a somewhat similar but fixed set of shunts across it, to slow it down with respect to $B2$ and to equalize operate and release time. The adjustment of the two potentiometers proceeds as follows. The time t_1 is set at 0, flux decay is chosen, and potentiometer 1 is adjusted to the point where the meter just starts to drop from the full reading. Then flux rise is chosen and the potentiometer 2 is adjusted to the point where the meter just starts to rise from zero. A fine check on these settings is to record the meter readings for a few equal small time steps and observe that the successive differences are consistent.

For applying an on and off battery condition as shown in Fig. 9 for the A contact, only one contact on one relay is used. The A contact circuit actually is made of two pairs of relays operated in a push-pull arrangement. The circuit may be changed to connect the four sets of transfer contacts into a lattice configuration to supply battery reversals to the apparatus under test. This is used in testing core materials to eliminate residual effects and for polar relays and similar structures.

The electronic control circuits which apply or stop the relay winding currents will be described later.

Timing Control System

The timing control system consists of a frequency source, a pulse forming circuit, three binary-quinary ring counters in tandem, a time selection circuit, coincidence circuits, memory circuits, and three control circuits, one for each of the A, $B1$–$B2$ and C relay circuits. A block diagram of the system is shown in Fig. 11.

Oscillator. The frequency source is a bridge stabilized oscillator. This is compared to the Bell System standard frequency for calibration. The accuracy of measurement, both for magnitude and the time scale depends directly upon this oscillator. The magnitude error enters through equation (5) where the cycle time T is exactly the interval for 1000 cycles of the oscillator frequency by virtue of the 1000 discrete steps in the three decades. The time scale is determined by these same steps. Hence the accuracy and stability of the oscillator enters directly; 0.1 per cent can be realized easily. This is as good as necessary as it exceeds the dc instrument accuracy.

Pulse Shaper—The pulse shaper is shown in Fig. 12. It consists of an initial cathode coupled amplifier followed by three more direct coupled stages. The square wave output of the third stage is differentiated by a small condenser for voltage spire production. It is a Schmidt[3] type circuit designed by I. E. Wood.

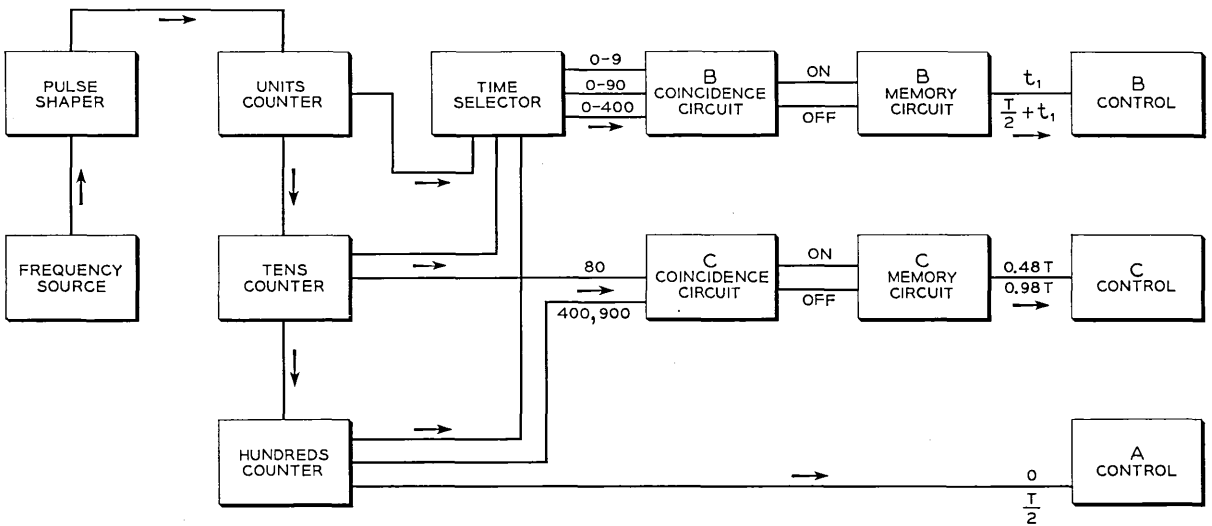


Fig. 11 — Block diagram of timing control system.

The differentiated output appears as a positive voltage impulse across the cathode resistor of a coupling tube, impressed on top of its normal dc bias due to current from the following quinary ring. The coupling circuit has the appearance of a cathode follower, but except during the short intervals of impulse transmission, the plate does not conduct. This lends itself to supplementary use of the rings as counter circuits. An on-off gate can be connected, as shown, to enable or disable the grid bias of this tube, without introducing any false starting or stopping counts.

Quinary-Binary Counters. The unit decade counter is shown in Fig. 13. It is a modification of Weissman's quinary-binary circuit[4] and was chosen because it provides simple two wire selection for the coincidence circuits. The principal modification consists of applying the shift pulses to the odd cathode lead rather than to a grid multiple, and using a direct coupled positive impulse for shifting.

The shaded tubes are non-conducting at 0 time and can be set in this condition by momentarily opening and then closing the reset key. The plate current from the right half of the zero tube passes through the cathode resistor of the output tube in the pulse shaper described above, and biases both tubes, the latter beyond cutoff. The plate currents from the other four tubes pass through the second cathode resistor having $\frac{1}{4}$ the resistance, developing the same bias.

A positive cathode impulse applied by the pulse shaper shifts the (0) tube from right to left half conduction. In shifting, the negative pulse from the prior non-conducting half of the zero tube cuts off the conducting half of the (1) tube, causing it also to shift. The shift of the (1) tube puts an aiding backward pulse into the (0) tube and a forward pulse into the (2) tube to keep it unchanged. Successive cathode impulses continue the ring stepping, the ring closing on itself after the fifth count. Thus the ability to shift properly depends solely upon the impulse waveform from the pulse shaper, which can be controlled independently of the ring circuit.

At the beginning of the fifth impulse, a square negative impulse is passed by the conducting half of the twin coupling diode to the (5) binary tube, causing it to shift. On the next round, the other half of the diode conducts, restoring the (5) tube. The twin diode behaves as an infinite capacitance in a circuit having zero time constant.

For each of ten successive impulses there is a different configuration of conducting tubes, and two wires, one to the quinary ring and the other to the binary tube, can identify when a selected state is reached by a coincidence circuit noting that a low voltage is present on each wire.

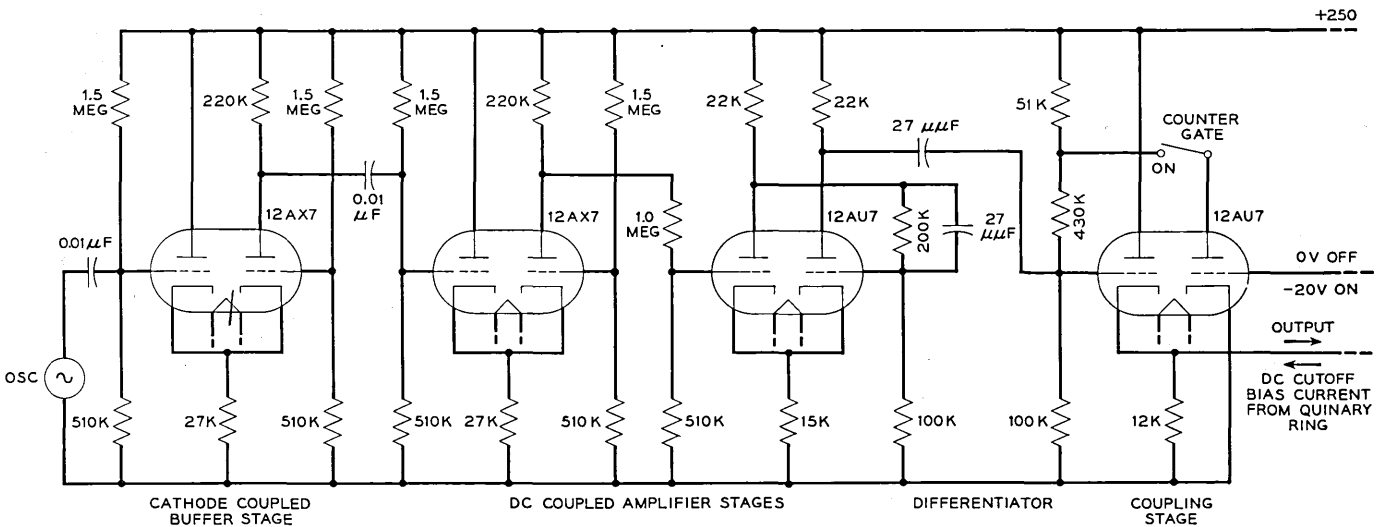


Fig. 12 — Schematic of pulse shaping circuit.

For the other nine configurations, one or the other of the wires will carry nearly the full battery voltage. The leads to the time selector switches are indicated by the arrows marked from 0 to 4, 5R and 5L, inclusive.

When the (5) tube reverts after the tenth impulse, its step plate voltage rise is differentiated by a small capacitor. This drives another pseudo cathode follower to shift the tens counter exactly as has been described for the units counter. For this tens ring, the take off leads to the time selector are identified as 00 to 50.

In a similar manner the hundreds counter leads are identified as 000 to 500.

Thus by exactly six wires, two for each decade, any cycle in 1000 can be selected, Note that for 0 and 500, needed for the *A* relay the (500) tube itself provides complete information. This eliminates the need for a coincidence and memory circuit for the *A* relay drive tube. Also by choosing 480 and 980 for the *C* relay, an abbreviated coincidence circuit can be used, as access to the units counter is not needed.

Neon lamps are provided as indicators for each ring to aid in circuit checking, trouble shooting, and as the counter indicator when used with the gate circuit.

The separation between successive time intervals which can be selected is one/one thousandths of one complete closure and open cycle, being 25 microseconds for a cycle time of 25 milliseconds, 100 microseconds for a cycle time of 100 milliseconds, etc. This is controlled by the discrete states of the counting rings used to generate the switching signals. This relation between the cycle time and the successive available time intervals is not a handicap because if the device is slow and a long cycle time has to be employed, it is slow because the flux buildup or decay is slow and hence closely spaced measurements are superfluous.

The maximum speed is 40 on-off cycles a second obtained with a 40-kc oscillator. The lowest speed is limited only by the ability of the vacuum tube filter to suppress instrument pointer vibration.

The counting ring system with its discrete steps precludes an automatically recording device as would be possible with a motor driven commutator and a gear driven take-off brush. However, the elimination of brush troubles is considered to be worth this sacrifice.

Time Selector. The time selection is controlled by two two-gang decade switches for the units and tens selections and one five position switch for the hundreds selection. The schematic is shown in Fig. 14. The dial positions are marked directly in time for a 10-kc oscillator, that is the units selector indicates one-tenth milliseconds, the tens milliseconds, and

the hundreds tens of milliseconds. For other oscillator frequencies the times indicated are scaled in proportion. Thus for a 40-kc oscillator, the indicated times are divided by 4 and the fine steps become 25 microseconds.

The start of the A cycle is marked by L500 and the stop by R500, the 500 tube itself serving as the memory circuit, and no coincidence circuit is needed.

The C cycle is wired permanently with a phase to close just before 500 and open just before 0. By choosing 480 and 980, access to the units decade is unnecessary and only four coincidence circuits are provided.

The start of the B cycle is marked by six coincidences, the end by five of these and the alternate connection to the 500 tube. The hundreds switch is simplified because the B relay always operates between 0 and 500 and releases between 500 and 1000. Therefore, the start and stop coincidence circuits can be wired permanently to the 500 tube and only

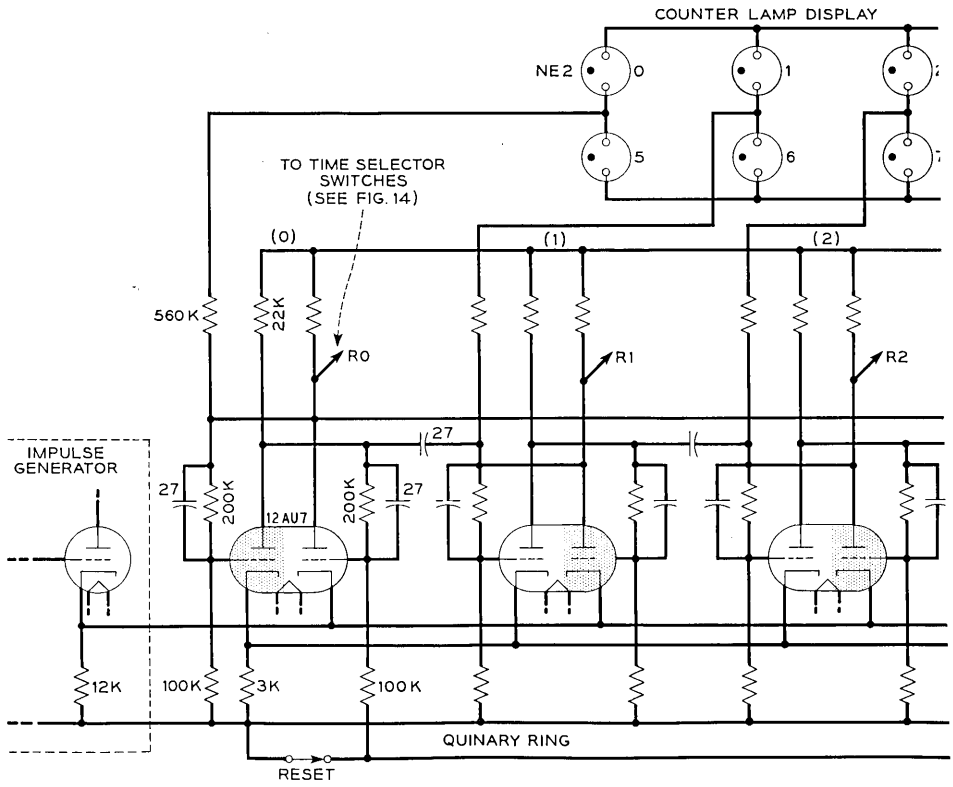
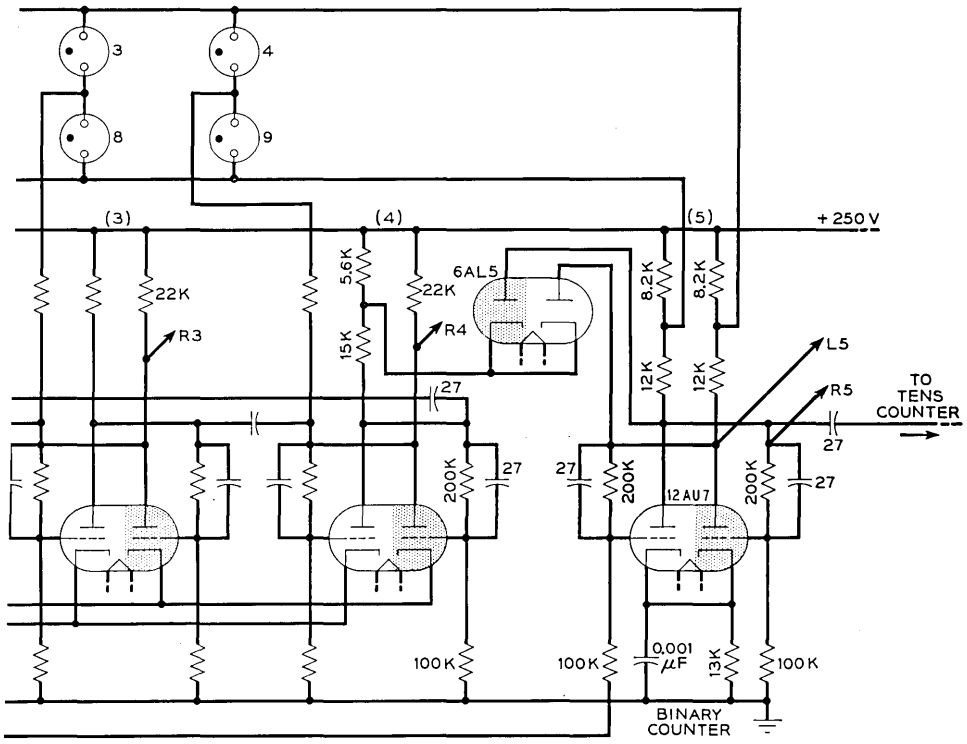


Fig. 13 —Interval

five switched wires are necessary. The change from flux rise to flux decay by another switch as shown in Fig. 10, converts the contact switching to provide the proper instrument closing time, with the time dials now reading time from the beginning of the open of the *A* circuit.

Coincidence Circuits. Each coincidence circuit consists of triodes, one for each marked wire. The circuit for the *B* cycle is shown in Fig. 15. Each selector switch wire connects to a grid, but all the cathodes of a set of coincidence tubes use a common cathode resistor. Except for the instant chosen, at least one coincidence tube has a high potential on its grid and the plate current through that tube holds the cathode resistor at a high potential. Only at the chosen time are all plates of the counter tubes conducting and therefore at a low potential. This abruptly drops the cathode potential of the coincidence circuit for one cycle of the oscillator, recurring every 1000 cycles. 500 cycles after each start pulse a similar stop



timer schematic.

pulse occurs in a duplicate set of coincidence tubes. These alternate pulses control the state of the memory circuit.

The C cycle coincidence circuits are similar except only four triodes are used for the start, and four for the stop pulses.

Memory and Relay Control Circuits. The memory and control circuits are shown in Fig. 16. The function of the bi-stable memory circuits is to accept the alternate start and stop pulses from the coincidence circuits, switch to the state representing the imposed condition, and hold it until the next successive imposed condition.

The function of the relay control circuits is to operate or release the relays under control of the memory circuits. The relay control circuits are direct coupled to the memory circuits and one or the other plate is cut-off by the grid bias condition imposed. The relays are in the plate circuits of the control tubes and either have full or no current applied. The operate and release times of the relays add a delay in the contact

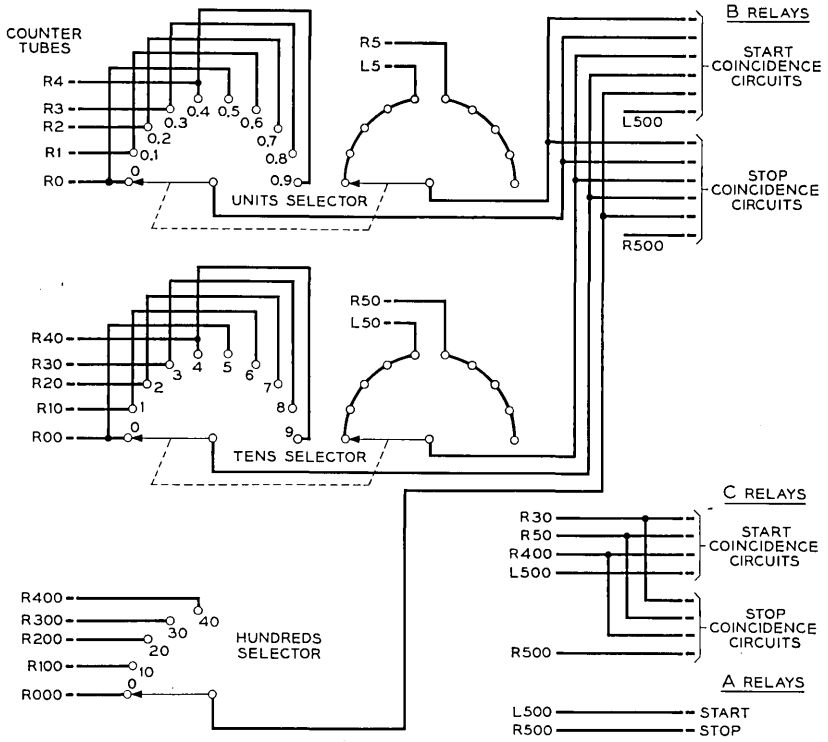


Fig. 14 — Time selector schematic.

actuation compared to the state of the memory tube. By making these small and equal, through selection of the relay design, the constant time lags do not alter the relative preset switching instants.

The relays are special mercury contact, single transfer, Western Electric 291-type. They have 38 gauge 1500 ohms, 14,600 turns windings and are adjusted to operate with 3.5 milliamperes and release with 2.0 milliamperes. The 5687 tubes switch from 0 current to 10 milliamperes through the relays. The release and operate times of the relays are about 1.3 milliseconds but as described earlier, are selectively adjusted to 1.5 milliseconds where necessary, by shunts such as those shown in Fig. 10.

The bi-stable memory circuits are switched at the grids through the twin diodes coupling the coincidence and memory circuits. While waiting for a shift pulse, the grid potential of the memory tube is lower than the cathode potential of the coincidence circuit, and the diode circuit therefore does not conduct. This is necessary because small voltage variations occur in the coincidence circuit cathode potential as the several tubes follow the decade counters. These small variations would shift the memory circuit falsely if directly connected. When the shift coincidence occurs, the cathode voltage drop is to a potential lower than that of the memory circuit grid and the conduction of the diode connects the two circuits. This negative impulse shifts the memory circuit. This drives the start grid to a new potential lower than that of the coincidence tubes cathodes, again cutting off the connection through the diode. When the coincidence cathode rises at the end of the shift pulse, the diode is even further cut off. The stop diode behaves in an exactly similar manner due to the symmetry of the circuits. The diodes thus afford a means of only momentarily making connections at the required instants, at all other times isolating the memory circuit from the remainder of the counting system.

Summary of Dynamic Fluxmeter Description

Up to this point, the circuit description has been concerned only with the dynamic fluxmeter. This circuit was designed and built first, and put into operation for dynamic current and flux studies of which one will be described later. About fifty standard type tubes in all are used, with good segregation of the circuit functions. This aids in localizing circuit troubles. Particularly useful is the neon indicator lamp display, also employed when the decade system, with the gate circuit, is adapted as a counter or time interval circuit.

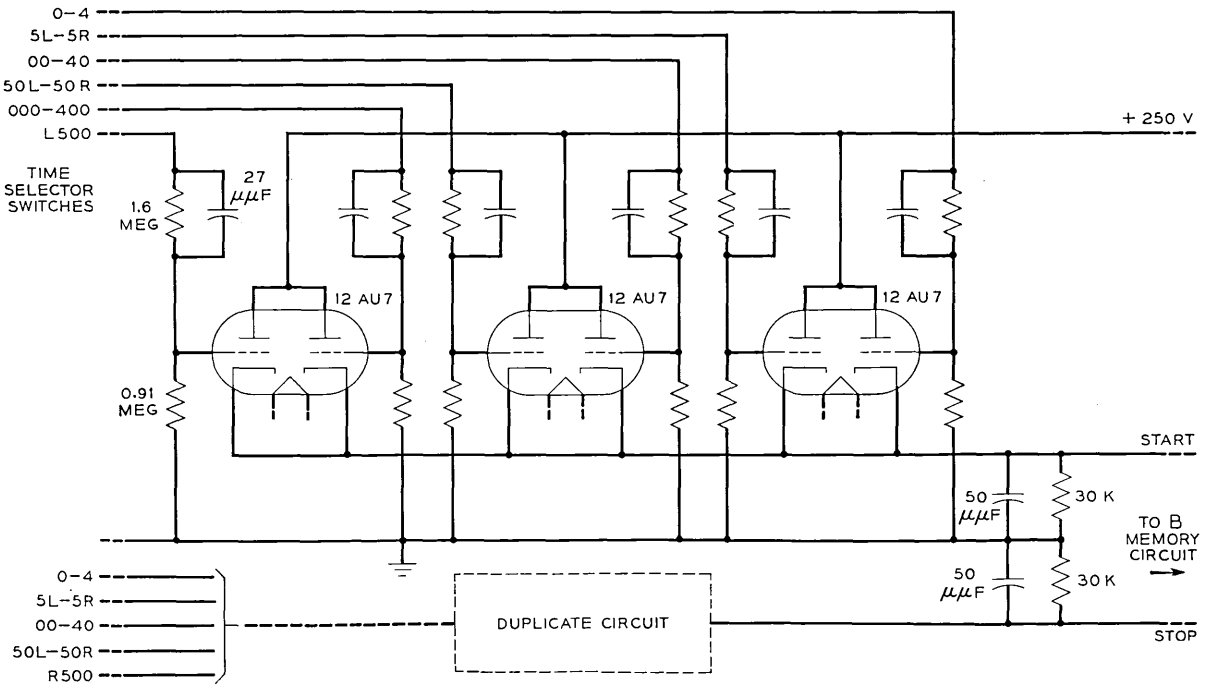
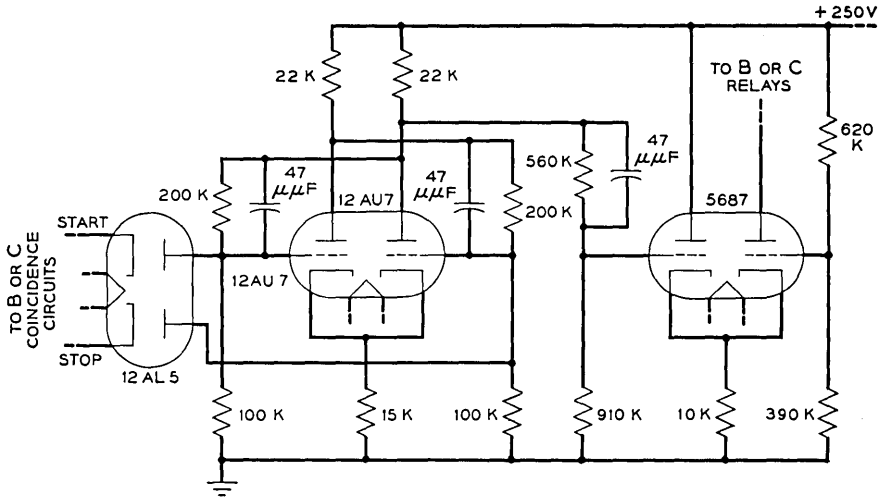
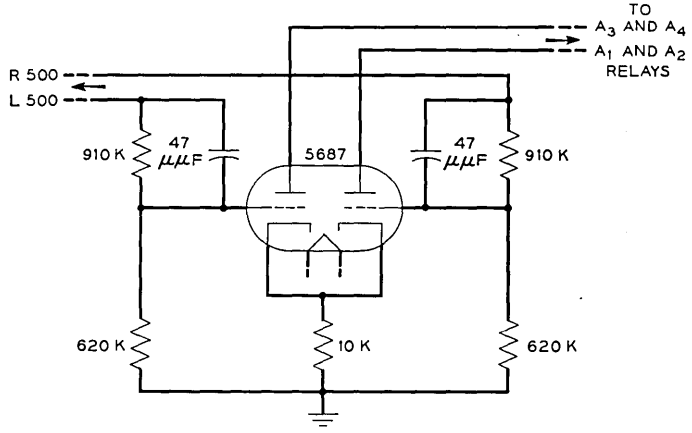


Fig. 15 — B-Cycle coincidence circuits.



B OR C MEMORY AND RELAY CONTROL CIRCUIT



A RELAY CONTROL CIRCUIT

Fig. 16 — Memory and relay control circuits.

Part II — Displacement and Velocity Measuring System

The Optical System[1]

A block diagram of the displacement and velocity measuring system was shown as Fig. 4. As part of this figure, a geometric schematic of the optical part of the system is shown. Most of the features of the optical system are due to the advice and assistance of associates in these laboratories who were connected with the sound motion picture development. The fundamental design is similar to the film reproducing system with certain necessary changes in dimensions.

Referring to Fig. 4, the elements are the lamp, condensing lens, a slit, objective lenses, a vane on the part whose motion is to be studied, and a photocell. The lamp and the condensing lens are the same as are used in the film reproducer. The slit is the same except that its width has been increased to 0.005″ and a feature added permitting its length to be adjusted from zero to about 0.1″. The objective lenses are inexpensive Bausch and Lomb achromatic lenses. The second objective lens is interchangeable for different focal lengths, and may be moved back and forth in the supporting tube for precisely focusing the image of the slit on the moving vane.

The three lenses and slit are mounted in a tube fastened to the lamp housing which in turn is supported by adjustable supports mounted on a vertical stand. To the lamp housing support is also fastened the photocell container with its first amplifier tube.

A permalloy shield surrounds the photocell except for an opening for access of light which passes by the moving vane. The shield is provided to prevent the changing stray magnetic field from the nearby electromagnet under test from affecting the photocell current. This system can be moved in any of three directions for lining up with the vanes of the device being tested. The relay or structure being measured is not fastened to the optical system support, but is secured to another appropriate stand resting on the same laboratory bench.

This system may seem more complicated than necessary at first sight, but its advantage is that it provides a rectangular beam of light of constant width and adjustable length at the point being measured. The necessity of numerous light shields and screens is done away with, although it is advisable to throw a black cloth over the whole apparatus when once adjusted to keep out the overhead illumination.

The reason for this last is of interest. The fluxmeter timing is controlled by a stable oscillator. The 60-cycle power for the overhead lights is also

well controlled. The two systems therefore operate in substantially a synchronized fashion when convenient cycle times are chosen, such as 10 cycles per second. The overhead fluorescent lamps fluctuate in light intensity at 120 cycles per second, so that the contacts in the fluxmeter operate in synchronism with the fluctuations in light intensity. Any measurements made with overhead illumination getting into the photocell will therefore have a superimposed 120-cycle ripple. For the same reason extreme care has to be taken to avoid any 60-cycle pickup in the apparatus. The lamp and the heaters on the first amplifier tubes are supplied with well filtered dc and cannot be operated by ac. Noise not in synchronism with the contacts is of little importance, as the averaging of the meter cancels it. A comment on the linearity of the photocell should be made at this point. This system operates on a variable width, rather than a variable density basis. Consequently the linearity of the system depends upon the uniformity of emission of the photocell surface. This can be verified by moving a vane with a micrometer and recording the output of the amplifier with a precision dc voltmeter. If the relationship is not linear other photocells can be substituted until sufficient linearity is achieved.

Amplifier System

General. Fig. 4 showed the three dc amplifiers in block diagram form. These amplifiers have each been designed to operate with full internal gain from dc to 10,000 cycles. The external transfer characteristic of each is controlled by its input and feedback networks. The design of these networks provides ideal transfer characteristics from dc up to 10,000 cycles. At 10 kc, the frequency response deviates by less than 3 db from the ideal. This same frequency corresponds to a time constant of 16 microseconds. On a small signal basis, the fidelity of measurement therefore will extend to events occurring in times of the order of 16 microseconds.

A more serious limitation to the accuracy is the finite plate voltage available for the output amplifier which results in amplifier over loading on large peaks. As will be shown, this can delay the response of the meter to sudden velocity discontinuities.

The amplifiers have been designed to operate from two voltage supplies, plus and minus 250 volts. The −250 volts is a series regulated three stage circuit with an output impedance of less than 0.8 ohm at all frequencies. It also serves as the reference voltage for a three stage shunt regulated +250-volt supply. Keeping the magnitudes equal minimizes errors due to power supply voltage variations.

The photocell amplifier converts the current from the high impedance photocell into a proportional voltage, having an internal impedance of 10 ohms.

The following amplifier-differentiator has an internal gain of 80 db from dc to 10 kc. When used as a differentiator the rising external gain characteristic reaches 67 db at 10 kc.

The output amplifier also has an internal gain of 80 db from dc to 10 kc. The feedback network includes an equalizer to produce current, and hence flux, in the air core output transformer proportional to input voltage from dc to 10 kc. The inductance of the transformer, with constant applied ac voltage, would cause a 6 db per octave decrease in current above the frequency where its Q is unity. This is counteracted by designing the external amplifier gain to increase at 6 db per octave, starting at the same frequency. The output amplifier thus has the same characteristics as a differentiator at high frequencies and its external gain is 64 db at 10 kc.

The overall gain of the system, then, is 131 db at 10 kc. Stability has been obtained, even with this much overall amplification without resort to compartmentation. Each tube stage utilizes shielded turret construction, exposed interstage leads are very short and only the input grid leads are shielded because they connect to controls on the front panel. A box shield surrounds the gain selector and the displacement-velocity switch of the middle amplifier, the lowest level point on the main panel.

This system uses an electronic differentiator. Ordinarily in analogue computers, these are avoided because of the high-frequency noise introduced as a result of the attendant large amplification. In the present system, the averaging of the succession of impulses by the dc instrument minimizes this effect. With each cycle a discontinuous section of the noise is sampled, containing a dc component, but as these are random in sign, their average gives rise to no error.

The system is dc coupled up to the output transformer. Slow drifts, due to grid currents or other reasons, do not result in instrument current errors because of the differentiating action of the transformer. No actual dc source appears in the meter circuit itself. The only concern here is to maintain the amplifiers somewhere near their best operating point. Actually the main source of dc drift is the temperature coefficient of the photocell and the voltage stability of the lamp power supply.

Calibration. The calibration of the system starts with a static measurement of the total displacement of the reciprocating motion to be studied. This can be done using thickness gauges or a tool maker's microscope. Then the device is brought into alignment with the light beam and cycled

by the A contact. The cycle time is chosen for complete operation and the time selector is set for a time near the end of the operate interval. At this time the tested device is known to be at its maximum displacement, and the instrument reading, using the displacement connection, corresponds. A convenient instrument indication is obtained by adjustment of the amplifier gain controls. For instance if the displacement were 0.040″, an instrument reading of 200 microamperes could be used. At any other time, when the parts are in relative motion, the instantaneous displacement is read directly from the instrument using the same scale conversion factor.

The calibration for velocity measurements depends upon the displacement calibration and relationship between the input differentiating capacitor and the resistor it replaces. Once a displacement gain setting has been chosen, it must not be altered during the associated velocity measurements except by calibrated steps. The differentiating capacitor has been chosen so that an instrument displacement deflection, corresponding to a given number of thousandths of an inch, represents the same number in inches per second. For the example given above, a 200-microampere instrument reading would represent an instantaneous velocity of 40 inches per second, with a linear calibration for intermediate readings.

A plot of measured displacement and velocity for a fast wire spring relay shown in Fig. 21, will be described when system errors are considered.

Photocell and Impedance Transformer Amplifier. A schematic of the photocell and dc impendance transformer is shown in Fig. 17. The high vacuum photocell has an impedance of thousands of megohms and acts essentially as a constant current device, the current depending upon the instantaneous illumination. This current is connected to the grid of a series stabilized[5] twin triode amplifier tube, to which grid also is connected the current through the feedback resistor and a balancing current from the −250-volt power supply. The zero adjustment is made under quiescent conditions for the desired dc output voltage, there being, of course, essentially no dc current in the grid.

The Western Electric low grid current 420A input tube is stabilized both for heater voltage and plate potential. It provides a voltage amplification of exactly $\mu/2$ or 35, by virtue of the upper tube having an impedance exactly equal to that of the plate of the lower tube. This input tube is mounted immediately behind the photocell in the same shield.

The output from the first amplifier tube is connected through a double shielded cable to the output tube mounted on the main chassis. The 5687 twin triode output tube uses both halves as cathode followers, one

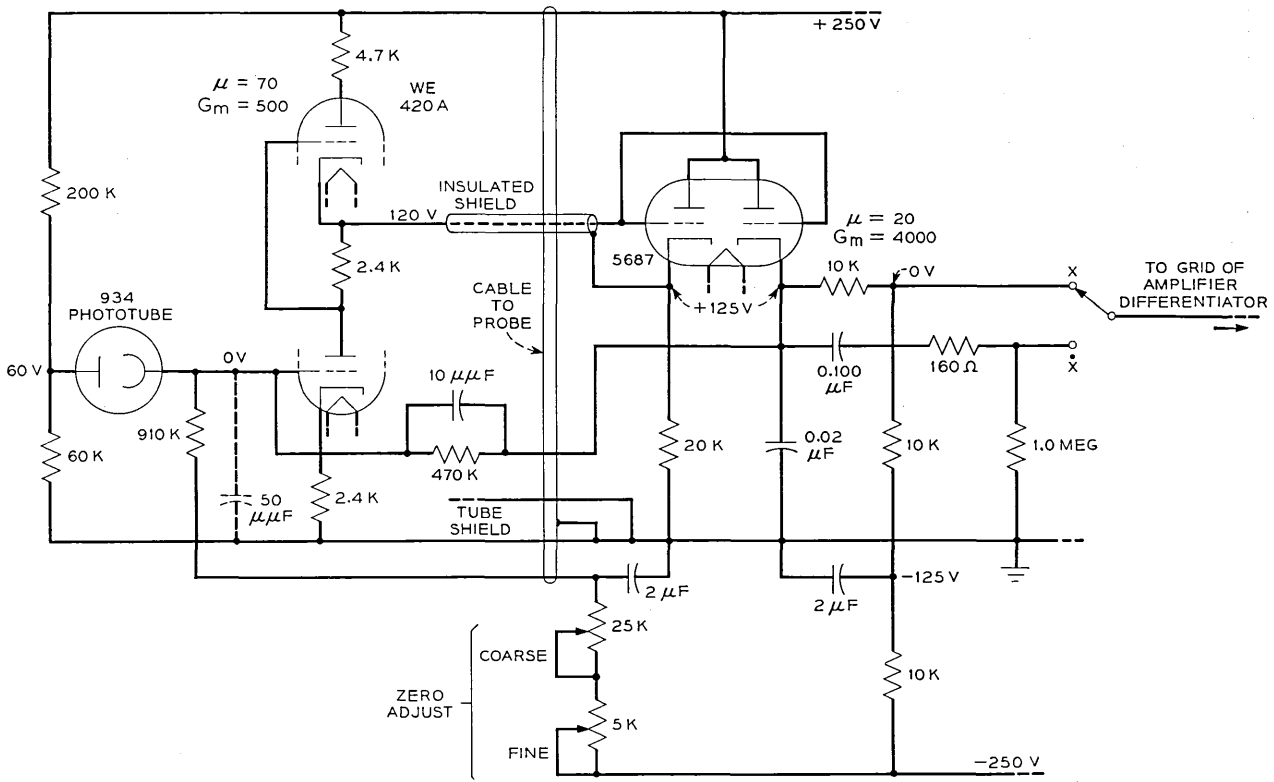


Fig. 17 — Schematic of photocell and DC impedance transformer.

to provide the output voltage and the other as a driver for the insulated inner shield to reduce the effective capacity to ground of the interstage cable connection. The cathode potential of the output tube is operated at exactly +125 volts, set by means of the zero adjustment previously described.

The input networks of the succeeding amplifier-differentiator are shown as part of this circuit as they provide the path for the output tube current and enter directly into the frequency response and loop gain cutoff design of this feedback amplifier. The dc grid voltage of the succeeding amplifier is 0, and this potential is found one-third the way down the cathode resistor which connects to the −250 volts. This point then provides the input to the following grid for displacement measurements and the input resistance for gain computations of the following amplifier is only this one-third part of the total cathode resistance or 10,000 ohms.

For velocity measurements, the succeeding grid is switched to the polystyrene differentiating capacitor whose grid side also is kept in readiness at 0 voltage by a grounded 1-megohm resistor.

For a change in current from the photocell due to a change in light, the feedback acts to supply an equal but opposing current from the output to keep the input grid nearly at its virtual ground potential. Thus the output voltage change E_s is given to a good approximation by the simple relation

$$E_s = \frac{I_s R_f \mu\beta}{1 - \mu\beta} \approx -I_s R_f , \tag{16}$$

where I_s = change in photocell current, R_f = feedback resistance, and $\mu\beta$ = loop gain.

A word about the differentiator connection is in order here. For this use, the cathode load approaches 160 ohms at high frequencies because the input grid of the next amplifier is a virtual ground point, and the load becomes equal to the phase shift controlling resistor in series with the 0.1 mf capacitor, the value of which will be discussed later. The impedance of the cathode follower without feedback is 250 ohms but this does not mean that it can be connected to a 250 ohm load and operated at the normal power output rating. The basic limitation for any tube is the allowable plate current change, regardless of the external load, consistent with never drawing grid current or being cutoff. For instance, in a cathode follower, if a load equal to 1/Gm is used, the small signal output voltage is only half of the applied grid voltage, a 6 db loss. Lower load resistances result in corresponding greater losses. In the present circuit, about an 8 db reduction in loop gain occurs at 10 kc

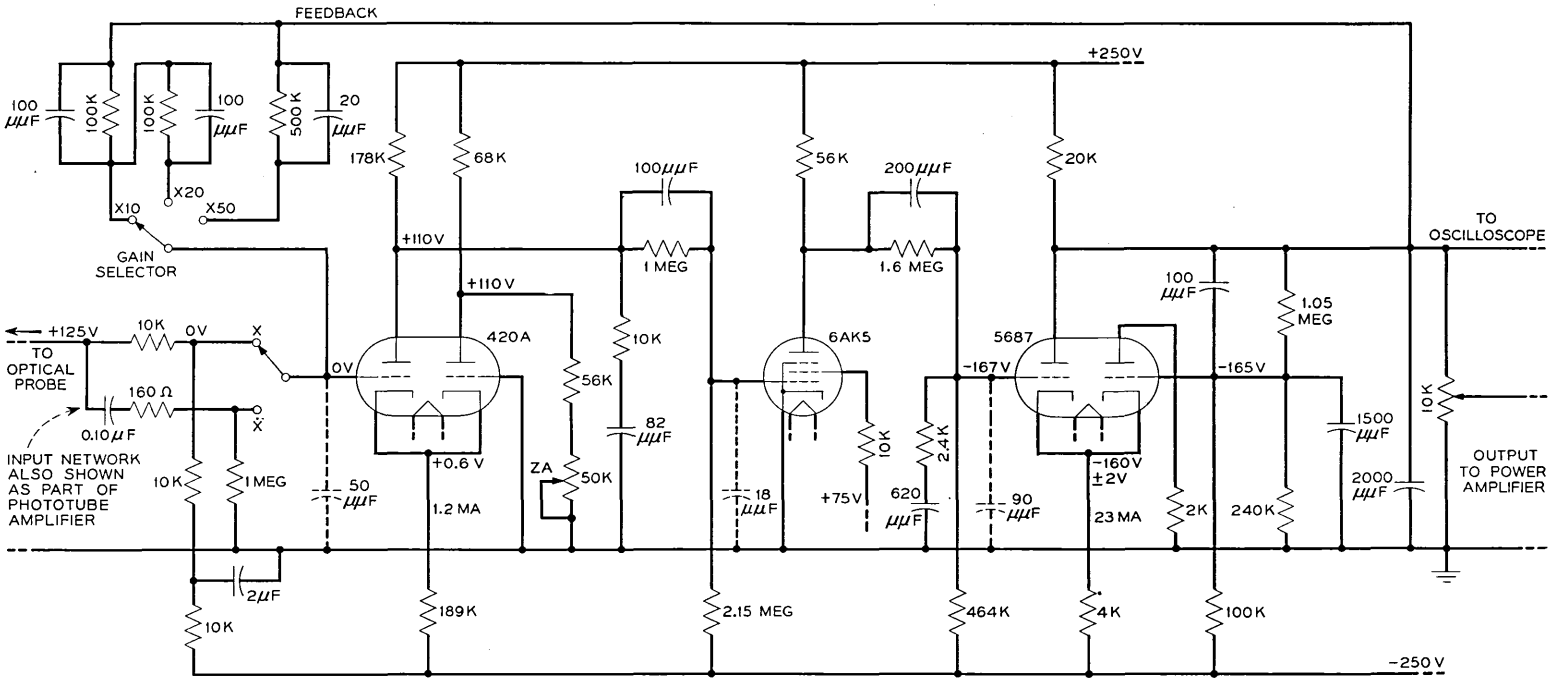


Fig. 18 — Amplifier-differentiator schematic.

leaving a net gain of about 22 db, including the 3 db effect of the input grid capacity.

Of equal concern is the reduced power handling capacity, but fortunately the energy in the frequency spectrum of the input signal diminishes rapidly with increasing frequency. A measure of this can be arrived at by the following analysis. The output voltage change for a motion of 0.040″ is about 0.1 volt. This total motion ordinarily never takes place in less than about 0.001 second, or a maximum rate of change of 100 volts per second. Now for a capacitive circuit

$$i = C \frac{dE}{dt} = 0.1 \times 10^{-6} \times 100 = 10 \text{ microamperes}, \qquad (17)$$

which is extremely small compared to the 12.5 milliamperes quiescent cathode current. Thus in choosing a vacuum tube to provide the desired amplifier output impedance characteristics, all other considerations have been cared for.

Amplifier-Differentiator. A schematic of the amplifier-differentiator is shown as Fig. 18. It has three stages, is direct coupled, and the quiescent output voltage is zero, set by the zero adjustment which controls the plate voltage of the inner tube of the input stage. The first stage is a conventional parallel stabilized circuit. It uses a high common cathode resistor to return the twin plate currents to the −250-volt supply. The first stage output is fractionated by an L pad to provide the proper dc bias for the second pentode stage. The output of the second stage is likewise fractionated for the dc bias of the final stage.

The third stage is a twin triode designed with an inside positive feedback loop. The two tubes have a common cathode resistor. After arbitrarily choosing one of the three grid resistors of the inside tube, the other two can be determined so that the proper quiescent bias and voltage changes result to make the plate current of the inner tube complement the plate current of the output triode itself. This maintains the total cathode current constant and establishes the cathodes as a virtual zero impedance to ground point. This realizes full gain from the output stage without another power supply. It also protects the triode from destruction if an accidental ground connection is made to the output.

The feedback circuit permits a choice of three values of resistances. The external amplification can be set at 10, 20, or 50. The maximum value of feedback resistance which can be used is determined by the value of the differentiating capacitor. It is set by the requirement that at least 10 db of loop gain remain at 10 kc.

It is now appropriate to discuss the scale factor for the differentiator

and to arrive at the value of capacitance to select. First we wish to measure reciprocating motions of the order of 0.050″ and actuating times of the order of 0.005 second. However, the motion takes place in the order of half the actuating time. Hence the average velocity will be of the order of 20 inches per second. An average numerical ratio of velocity to displacement then is 20 divided by 0.05, or 400. Remembering this is average and some part of the velocity will be higher the nearest decimal factor is 1000.

In this system, as a device being tested is a part of it, the instrument time scale necessarily is identical with real time, that is it cannot be "slowed down" for more accurate time settings, as can analogue computers.

Now once a convenient gain setting for displacement measurements has been determined, if a capacitor were substituted for the amplifier input resister, related by the equation $RC = 1$ second, C would be 100 *mf*, as $R = 10{,}000$ ohms. This would be a substitution where the scale reading for a given number of mil-inches would represent the same number of mil-inches per second. Fortunately, however, we wish rather to represent a much higher velocity of inches per second. This results if the capacitor is reduced by the same factor of 1000. The desired capacitance thus has a value of 0.10 *mf*. The type used is a stable, low soak polystyrene dielectric one having ½ per cent accuracy. In this amplifier the feedback and input resistors also must be of ½ per cent accuracy as upon these depend the calibration. The 10,000 ohms input resistor necessarily is a non-inductive wire wound type with a 10 watt rating, for stable performance with the actual 1.5 watts. The feedback resistors are of the deposited carbon type.

The remainder of the design now follows. To limit the differentiation action to a 10-kc range, a resistor in series with the 0.1-*mf* capacitor of 160 ohms is used. A transfer gain of 70 db, which is 10 db less than the internal gain of 80 db, is a voltage ratio of 3,160. Multiplying this by the input resistor value of 160 ohms yields 500,000 ohms as the maximum feedback resistance. Then as an amplifier with an input resistor of 10,000 ohms, the gain is 50. Resistances for lesser gains of 20 and 10 are scaled in proportion.

The linear output voltage range for this amplifier is ± 40 volts. This is much more than needed. For displacement measurements the maximum swing is about 5 volts. In the discussion, we found a maximum capacitor current of about 10 microamperes which also results in an output swing of 5 volts, during velocity measurements.

Output Amplifier. The schematic of the output amplifier is shown in Fig. 19. This amplifier also has three stages, the first two of which are

like those of the amplifier just described. Similar considerations for the overall design were used here and the loop gain cutoff control uses the same principles, modified to take account of the inductor load and its constant current feedback equilization.

The output tube is a 6L6 with an 874 voltage regulator tube in the cathode return to the −250 volts. The 6L6 is operated at 40 milliamperes, which is within the current rating of the 874. This stage was designed for the widest possible output voltage swing and for plate current cutoff independent of plate voltage, both characteristic of pentodes.

The transformer has a 900-cycle self resonance with its equivalent shunt capacitance. At higher frequencies it behaves as a capacitor as far as the tube and loop gain are concerned. Its Q is unity at 56 cycles.

The function of the output amplifier is to convert the input voltage signal into proportional current in the inductor. Constant voltage gain would not do this. For frequencies above 56 cycles, the current would drop at a 6 db per octave rate with increasing frequency, because of the reactance.

Note that if the voltage across the inductor rises with increasing frequency, then the current through the winding will be proportional to a constant input voltage regardless of the equivalent shunt capacitor. The capacitor will draw greatly increased current but as long as it is uniformly distributed, it will not affect the winding current itself. If series output feedback were used, the two currents could not be separated and a more complex equalizer would be necessary to provide the necessary frequency response. The choice of the value of inductance for the output transformer will be discussed after errors in the system have been considered.

The rising gain is produced by shaping the feedback, starting an insertion gain rise at 56 cycles with the 0.057-*mf* capacitor, and continuing to 10 kc. At this frequency the rising characteristic is discontinued by the 280-ohm series resistor, and the capacitors shunting the feedback resistors. Over the rising characteristic range the equilization introduces 90° adverse loop phase shift and above 900 cycles the output transformer adds another 90° phase shift. Thus from 900 cycles to 10 kc the amplifier just skirts being Nyquist stable.[6]

The internal gain of this amplifier is 80 db. At dc the external gain is 21 db. It rises to 67 db at 10 kc because of the feedback equilization.

Output Amplifier Square Wave Response

The square wave response of the output amplifier is shown in Fig. 20. Two effects are evident (1) a finite time for the major change to occur,

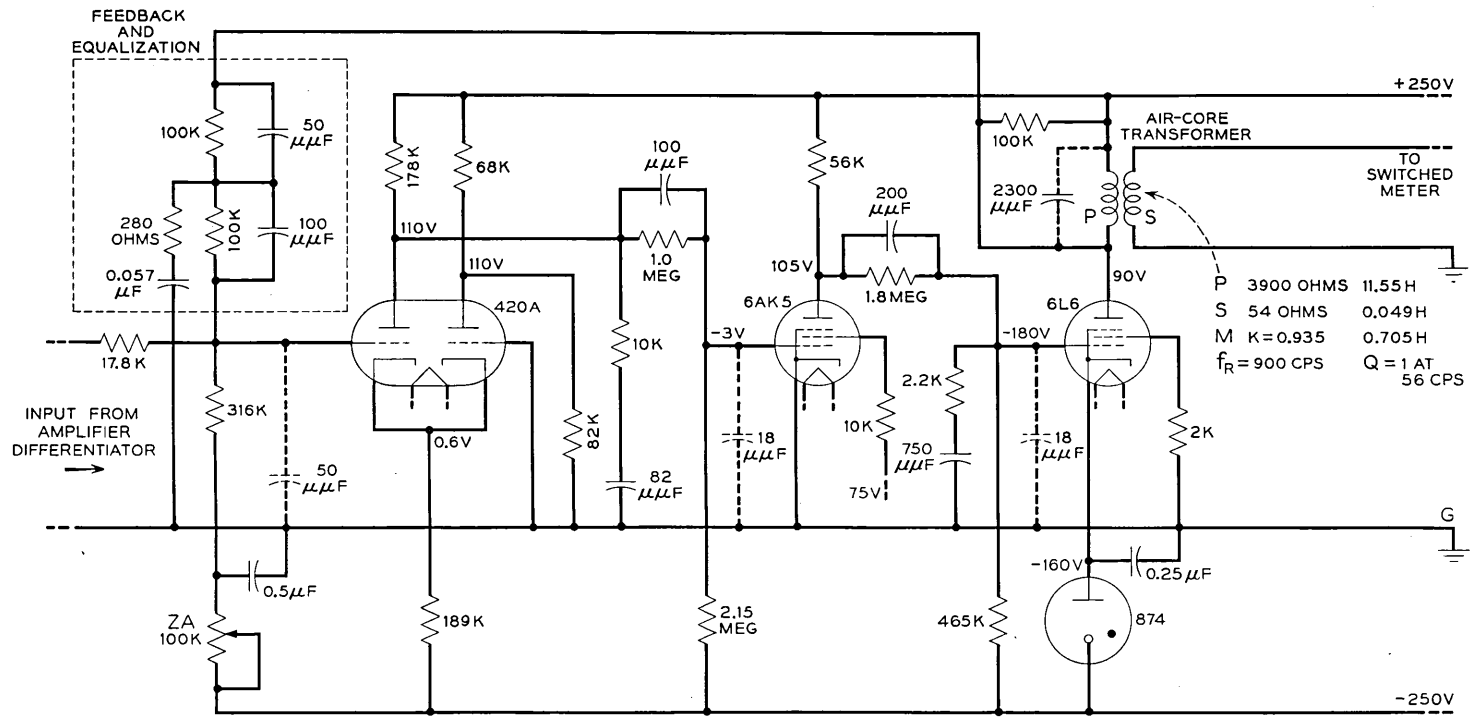


Fig. 19 — Output amplifier schematic.

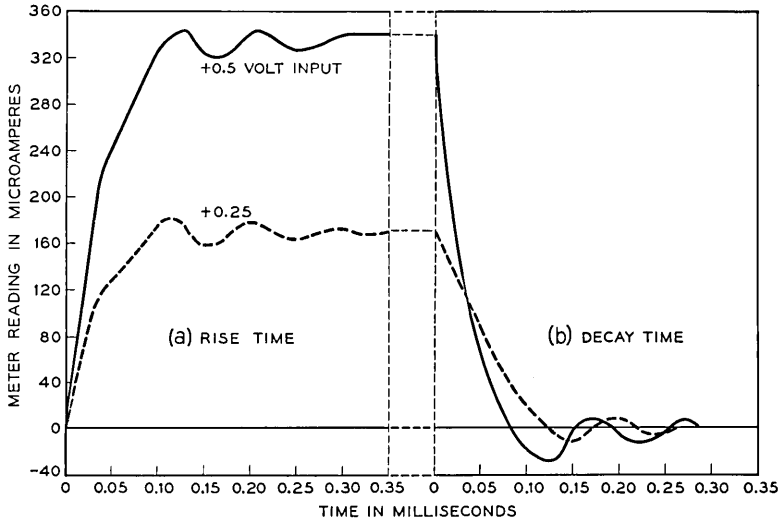


Fig. 20 — Output amplifier square wave response.

followed by (2) an oscillation caused by the frequency response of the amplifier deviating from the ideal for frequencies above 10 kc. The former effect is caused by the finite power supply voltages available in the final stage of the output amplifier.

The current in the primary of the output transformer represents the function being measured. If a step discontinuity occurs, as during velocity measurements, then the current suddenly has to be changed to a different value. The rate of change of an increase in current which can be developed is proportional to the power supply voltage, and hence is finite. For a decrease, the distributed capacitance of the winding and the shunting resistors delay the current decay. These result in a delayed transition from one condition to the other, occuring, from Fig. 20, in about 0.1 millisecond.

A discussion of the operating conditions of the system leads to a method of evaluating these forms of error and determining what limitations are imposed upon the use of the measured data. It will be shown that only immediately following velocity discontinuities are the data not usable. Displacement data are never in difficulty from an overload standpoint and fortunately we ordinarily are not too interested in velocities after impacts.

Discontinuity Errors Due to Overloading

The arbitrary gain setting for the output amplifier is chosen to provide about the full scale of 200 microamperes to represent the full displace-

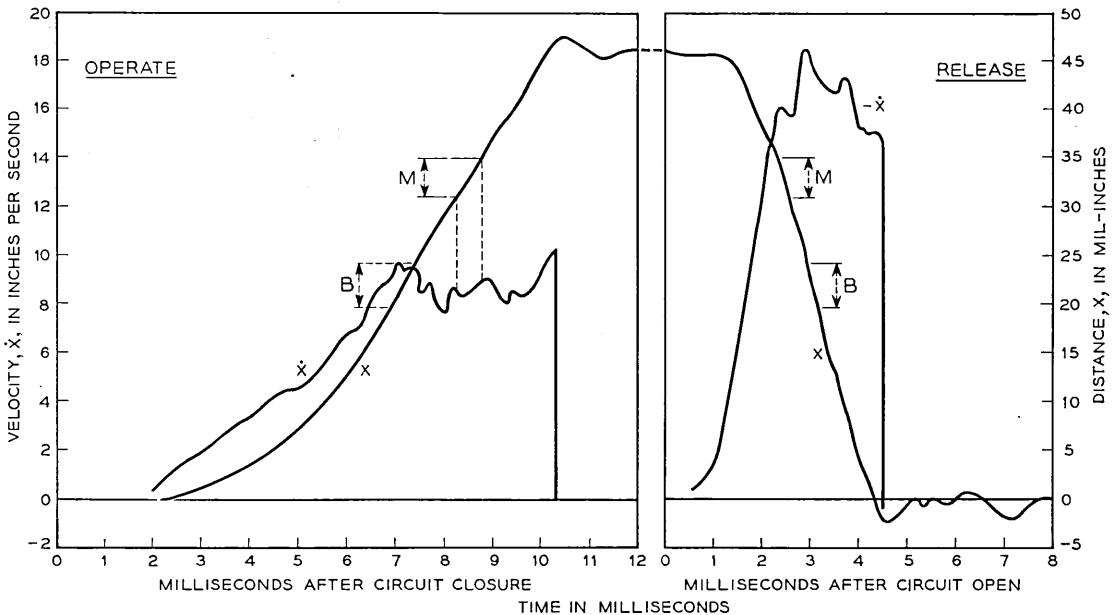


Fig. 21 — Displacement and velocity measurements of the AF-type relay. M — on operate, make contacts closed in this region, and on release, make contacts open in this region. B — on operate, break contacts open in this region, and on release, break contacts close in this region.

ment. The output transformer obeys equation (8), and its characteristics are tabulated on Figure 19. From this equation and a cycle time of 0.1 second we find that the change in plate current for a 0.705 henry mutual inductance amounts to 2.83 ma. The dc plate voltage change therefore is 11.32 volts and the corresponding input dc voltage is 1.0. These values are extremely small compared to the voltage swings available, but nevertheless the fact that the instantaneous voltage swings are finite imposes a limit to measurement of step discontinuities.

One direction of transformer current change is where we wish to decrease the current suddenly. The signal blocks the output tube and the transformer current decays through the equivalent of a 50,000-ohm shunt. The primary inductance is 11.5 henries and the time constant therefore is 0.23 millisecond. The quiescent current is 40 microamperes, so a drop of 2.8 microamperes, assuming the 6L6 blocks completely, would require at least 0.02 millisecond. This means that if there is a measurement discontinuity, the best the circuit can do is to respond as the first part of an exponential over at least this time interval, rather than abruptly. This is shown in Fig. 20, where the indicated time actually approaches 0.1 millisecond. This extension evidently is contributed to by the distributed capacitance of the winding.

The current rise case is about as favorable. The current rise rate is limited by Lenz's law to a maximum value of

$$\frac{di}{dt} = \frac{E}{L} = \frac{265V}{11.5H} = 23 \text{ ma/ms.} \tag{18}$$

A change of 2.83 microamperes therefore requires at least 0.12 millisecond. This is also shown in Fig. 20.

The foregoing discussion has been directed toward establishing that the initial delays of Fig. 20 are entirely overload effects and not a frequency response effect. Thus if required rates of change during a test do not exceed the available rate, then no error of this type will occur.

Fig. 21 is a measurement of a fast relay. One curve designated x, is the displacement versus time. The other curve marked $\dot{x}$, is the velocity. In neither the displacement curve nor the velocity curve before impact, do the rates approach the overload condition.

The impacts are shown on the velocity curve to have the extreme rate of change. Such abrupt curves cannot be taken as being literally true. Following such abrupt changes the 10-kc oscillations also are found, but have not been drawn as they have to be discarded.

Discontinuity Errors Due to Averaging

Another effect acts to limit the sharpness of measurement when discontinuities occur. This is due to variations in successive operations of the device being tested. Obviously, if the actuation time varies over a range of $\pm$ 0.1 milliseconds, then any device such as the present one which measures by averaging many measurements, will not measure accurately in the close vicinity of such a discontinuity.

For a displacement measurement made at the average time of armature impact, part of the measurements will be before impact, the others after. The meter therefore will register too small a displacement. For another measurement a short time earlier or later than the impact part will be before and part after the average instantaneous displacement but the instrument average will still be good because the function now is smooth. The only displacement error then is just at impact and when the data are plotted, appears as a rounding of the curve. A good correction can be made merely by continuing the adjacent slopes to a sharp intersection.

A more striking effect is observed for velocity. At the moment of armature impact against core, the velocity drops suddenly from maximum to zero. The indication shown by the instrument then for this averaging type error is just half the final velocity. Another measurement made a short time earlier is good and the slope from that region can be extrapolated to the instant of the half velocity indication as the plot for the true velocity.

Choice of Output Transformer Inductance

The output transformer primary inductance is a compromise between output amplifier gain limitation due to equivalent shunt capacity of the inductor and the lowest desired cycle time. It was shown above that at at 0.1 second cycle time only 1 volt is needed for a full scale deflection whereas 5 volts is available. As the instrument current is inversely proportional to the cycle time, 1.0-second cycle time will provide about half scale. The 11.5H chosen then, just barely covers the timing range of present interest. The impedance of the equivalent shunt capacitor at 10 kc is approximately equal in magnitude to the transformer dc resistance, and full internal gain is available over the operating frequency band.

The delay time is not altered by a change in inductance. For instance, if the inductance were halved the rate of current rise would be doubled but twice the change would be necessary for the same instrument deflection. The same result can be had merely by reducing the gain by one-half.

If longer cycle times were needed without the short times of present interest, the inductance could be appropriately increased, and the amplifier bandwidths correspondingly decreased, without loss of accuracy.

Summary of Amplifier Discussion

The foregoing discussion of the amplifiers covers the design considerations, the required frequency responses, and errors because of divergence from the ideal. For displacement measurements, all events occur in intervals greater than the minimum which the amplifiers can follow. This is also true of velocity measurements as regards the photocell amplifier and differentiator. The output amplifier does overload momentarily at jump discontinuities of velocity, and in the time vicinity following such events is in error.

Another type of error caused by variations in successive operations of the device being tested can be corrected by use of the measured rates just before and after the discontinuity.

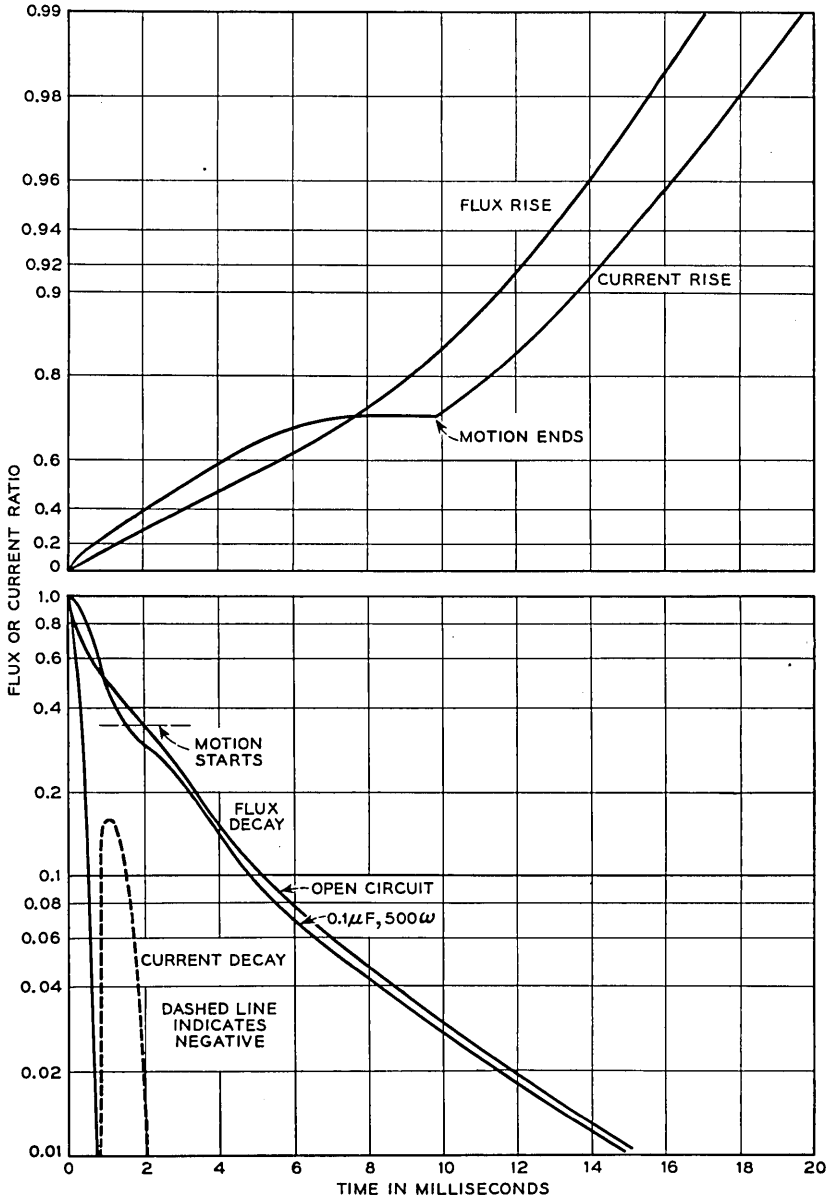


Fig. 22 — Winding current and flux measurements of an AF-type relay.

Part III — Applications

NEED FOR DYNAMIC FLUX MEASUREMENTS

The results of an experimental determination of dynamic flux rise and decay in solid core electromagnets are of general interest. Analytical solutions have been obtained for linear infinitely long rods or toroidal shaped structures where geometric simplicity exists. These solutions are in infinite series form. These solutions show that an elementary perfect representation can not be expected, even for these simple cases. Furthermore, an electromagnet has other factors which make any attempt at an analytic solution impractical. Some of these are:

(a) The magnetic material is non-linear.

(b) The flux density is non-uniform because of leakage flux.

(c) The varying geometry of the magnetic parts, including the necessary gaps, make the boundary value problems unmanageable.

(d) Motion of the armature during operate and release.

These all lead to the conclusion that a quick and accurate method of measuring dynamic flux changes is necessary for fundamental studies of the dynamic behavior of electromagnets.

As an example, Fig. 22 shows dynamic flux and current rise and decay measurements made on the same relay used for the dynamic motion studies. These data are plotted on semi-log graph paper as on such a plot an exponential curve becomes a straight line. The current rise curve shows the dip due to armature motion, ending abruptly when the motion is completed. The flux rise curve starts off nearly as an exponential but rises more rapidly after armature motion starts.

Two decay curves are shown, one with an open circuit and one with a contact protection network. Associated with the latter is the winding current which flows through the network. In the open circuit case, even without the effect of armature motion, the flux decay is not exponential. The flux decay with the network has a somewhat oscillatory shape about the open circuit curve. The current itself does complete one heavily damped cycle. The reversed current flow is shown as a dashed curve.

The decrease in flux after a short interval compared to the open circuit case, demonstrates that such a network can decrease the release time as well as afford contact protection.

For more fundamental studies, it is better to study flux behavior with the armature held fixed and avoid the motional effects.

DYNAMIC FLUX DECAY AND DEFINITION OF EQUIVALENT CORE CONDUCTANCE

Fig. 23 shows another measured flux decay curve with the armature locked in the operated position. Also shown are two one-term exponential

equations for comparison. For such an open circuit decay curve, the flux does not abruptly drop to zero because of induced core eddy currents. The analysis of these data thus partly resolves into the determination of some method for conveniently representing the eddy current effects.

NEW APPROXIMATE FLUX DECAY EQUATION

For a first approximation, use can be made of linear circuit theory. Assuming the core behaves as a coupled single turn of conductance G_e and inductance L_1, and defining the core time constant as:

$$t_e = L_1 G_e, \tag{18}$$

the flux decay equation is the well known solution:

$$\frac{\varphi}{\Phi} = e^{-t/t_e} \qquad t \geqslant 0. \tag{19}$$

It is clear that the dynamic flux decay cannot be represented by the above exponential equation, which has been drawn on Fig. 23 as the upper straight line.

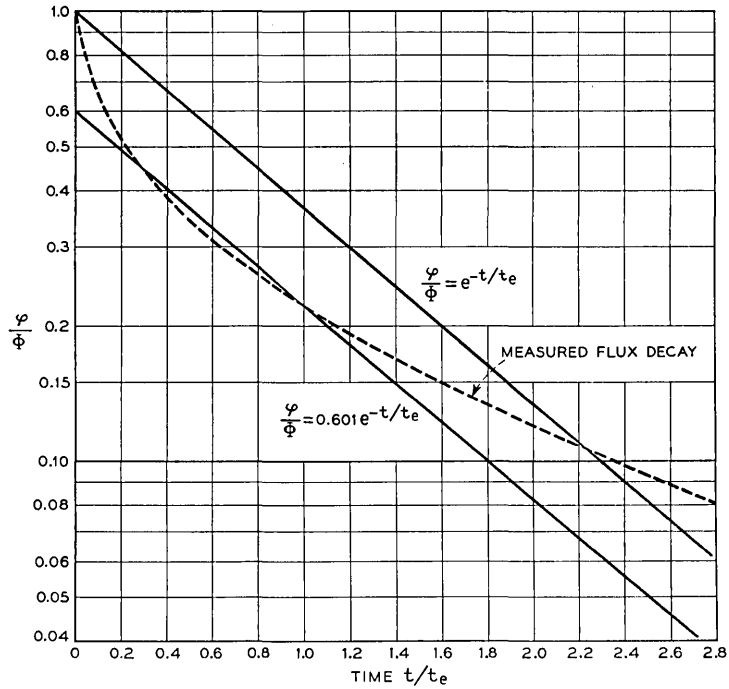


Fig. 23 — Comparison of measured open-circuit flux decay with one-term exponential equations.

The true decay curve has an initial steep slope followed by continuous curvature. A better approximation is to stipulate an initial jump discontinuity. A good fit to the flux range in which release of electromagnets occurs, shown by the lower straight line, is

$$\frac{\varphi}{\Phi} = .601e^{-t/t_e} \qquad t \geqslant 0. \tag{20}$$

This discontinuity of flux in the first approximation is just the reverse of the continuity of flux concept usually used. Of course, no actual discontinuity occurs, as shown by the true decay curve. The failure of any single exponential equation to represent the true curve merely makes clear the fact that the behavior of the core is not that of a single coupled turn, but rather is that of an infinite line. Bozorth[7] gives the intercept as 0.691 for the linear case, which does not represent the continuously curving decay which actually occurs.

The chosen intercept of the first approximation curve at $t = 0$ is admittedly somewhat arbitrary. It was arrived at in a broader study including flux rise curves. Accepting this equation, a convenient determination of t_e can be made similar to that for exponentials. If t is set equal to t_e, then

$$\left.\frac{\varphi}{\Phi}\right|_{t=t_e} = 0.221. \tag{21}$$

Thus, after measuring a dynamic flux decay curve, the time can be determined for which the above ratio obtains. This directly is t_e. From linear circuit theory and Lenz's Law, the inductance for one turn is:

$$L_1 = \frac{\Phi}{NI}, \tag{22}$$

whence G_e can be determined.

Now because of magnetic saturation and the shape of the hysteresis loop, the values will depend upon the particular final ampere turns (NI) used in the experiment. For uniqueness and uniformity in rating electromagnets for comparison purposes, the particular set chosen is that for which L_1 is a maximum. For comprehensive operating studies, measurements of course have to be made under the actual conditions of interest. Except for rating purposes, t_e is not ordinarily split up into components.

Thus, while the rated value of G_e for a particular electromagnet has the dimensions of conductance, it includes other factors as well. Some are: (a) Core material conductivity, (b) Magnetic non-linearity, (c) Shape of the hysteresis loop, (d) Non-uniform flux distribution, (e) Effect of pole

face and other air gaps, and (f) Decay of leakage field. It, therefore, also has the nature of a mop-up factor in which is included other effects not explicitly covered in the elementary analysis.

The above discussion has been directed toward establishing an easily determined core time constant having characteristics useful for rating purposes, and demonstrating that the distributed nature of the core eddy currents precludes an accurate representation of the core as a single coupled turn.

Open circuit flux decay measurements of round and rectangular solid core electromagnets of magnetic iron, 1 per cent silicon iron, and 45 per cent permalloy all are accurately represented by the measured flux decay curve of Fig. 23, after establishing appropriate values for G_e and L_1 in each case.

DYNAMIC FLUX RISE

Dynamic flux rise curves, measured with the relay armatures blocked open, are shown in Fig. 24. A family of curves, is needed because of the added variable of the winding. The value of t_e is determined by the method just described, except that the armature is held open, resulting primarily in a new and lower value of L_1. The winding may be characterized by its coil constant:

$$G_c = N^2/R,$$

where N = number of turns, and R = dc resistance of winding circuit. These curves are a composite of measurements on many structures and fit all data to within a few per cent.

An empirical closed form expression has been determined which fits these data, and in fact was used to compute these curves. However, once the curves have been plotted there is no further need for the rather complicated expression. Any particular rise curve desired can be interpolated from the drawing.

These curves again show, with windings of relatively low coil constants, a divergence of the dynamic flux rise from being a straight line. The particular curve $G_c/G_e = 0$ actually is a decay curve because such a rise measurement would involve the use of both an infinite voltage battery and an infinite resistance winding. Now for this case, as well as for all the others, the rise and decay curves differ because of the shape of the hysteresis loop, the decay curve persisting longer. This effect is most evident for the last few per cent of the flux change. However, the differences are small on a full range basis. The curves shown near $G_c = 0$ are

a compromise for this effect. Actually the rise curves do not cross, but they do approach each other.

Coil to core constant ratios now in use range from 2 upwards, the case of 2 being the best condition for 25 watts power. An examination of this particular curve shows some curvature for small times. This effect of an initial increased rate of rise reduces the operate time of an electromagnet, and is automatically taken advantage of in the experimental design of windings.

With a winding as part of the system, an elementary consideration shows why this curvature exists. This follows from

$$N\frac{d\varphi}{dt} = E - iR, \tag{23}$$

where: N = number of winding turns, E = applied battery voltage, R = resistance of winding, i = instantaneous current, and φ = instantaneous flux. This equation is exact and holds whether or not there are eddy currents. At the instant of circuit closure there are no winding or eddy currents (because of leakage inductances if for no other reason) and the initial rate of flux rise is dependent only upon the number of turns and the applied voltage. After a transition interval, the eddy currents become effective and slow down the rate of flux rise. For any given total flux, the time required, therefore, is less than that given by the well known equation:

$$\frac{\varphi}{\Phi} = 1 - e^{-t/L_1(G_e+G_c)}. \tag{24}$$

However, for large coil constants where G_c predominates the effect diminishes and the above equation is an excellent representation and desirable for its simplicity.

NEW FIRST APPROXIMATION FLUX RISE EQUATION

Returning to Fig. 24, while the effect we have been discussing is all important for open circuit flux decay, practical coil constants at present do not have ratios to the core constant much below 2. These curves do not diverge greatly from straight lines on this plot. Can a minor correction term be made a part of the simple equation heretofore used which will retain its simplicity and extend its accuracy to windings now used?

One such equation is

$$\frac{\varphi}{\Phi} = 1 - e^{-t/L_1(G_c+G_e e^{-G_e/G_c})}, \tag{25}$$

where Φ is the final flux. This may also be written in integral form as

$$t = (G_c + G_e e^{-G_e/G_c}) \int_0^{\varphi} \frac{d\varphi}{NI - Ni}. \tag{26}$$

The change is the introduction of the exponential modifying the core constant G_e.

Fig. 25 shows comparisons of curves for four different coil to core constant ratios. In each case, the actual flux rise, the older equation, and the new first approximation are shown.

For large coils (as an example on Fig. 25, $G_c/G_e = 10$) the actual flux rise is well represented by either expression, their accuracy being within 1 per cent.

For very small coils (on Fig. 25, $G_c/G_e = 0.5$) it is clear that the older representation is never a good approximation of actual flux rise. The new approximation represents the start of the dynamic flux rise quite well, but is not accurate for the second half. However no single exponential

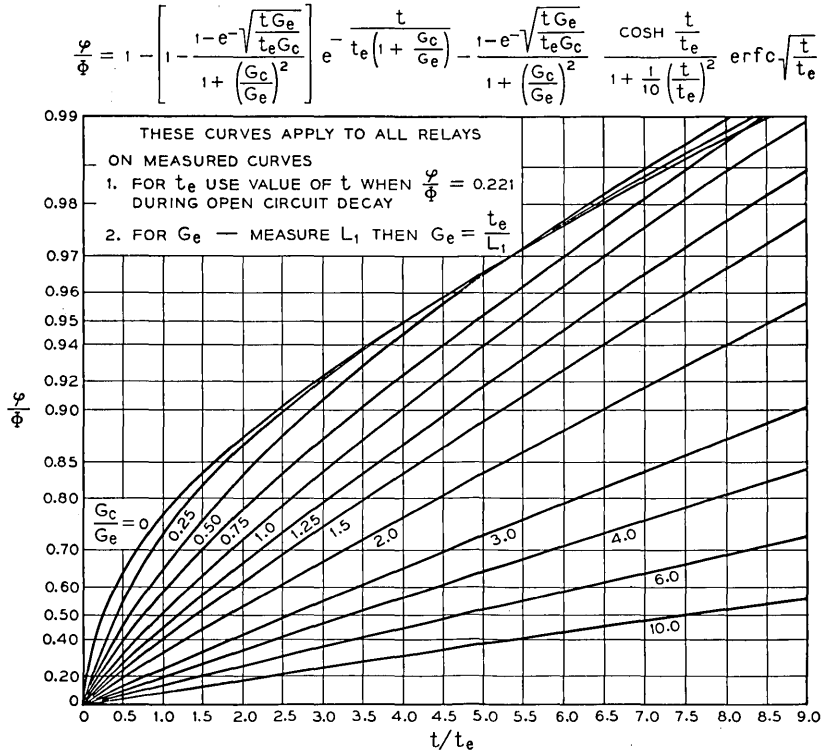


Fig. 24 — Flux rise curves.

equation can represent the actual rise when the plot is as curved as this one is. The best that can be done is to approximate the more important initial part with a straight line, and in this sense the new approximation shows a good fit.

A typical speed relay will have a value of G_c/G_e around 2 and the range from 1 to 10 covers all relays in this class. For a ratio of 2, the new representation of flux rise is within 2 per cent of the actual rise, as opposed to about 5 per cent with the earlier equation. For the ratio of 1, the accuracies are 4 per cent and 11 per cent respectively.

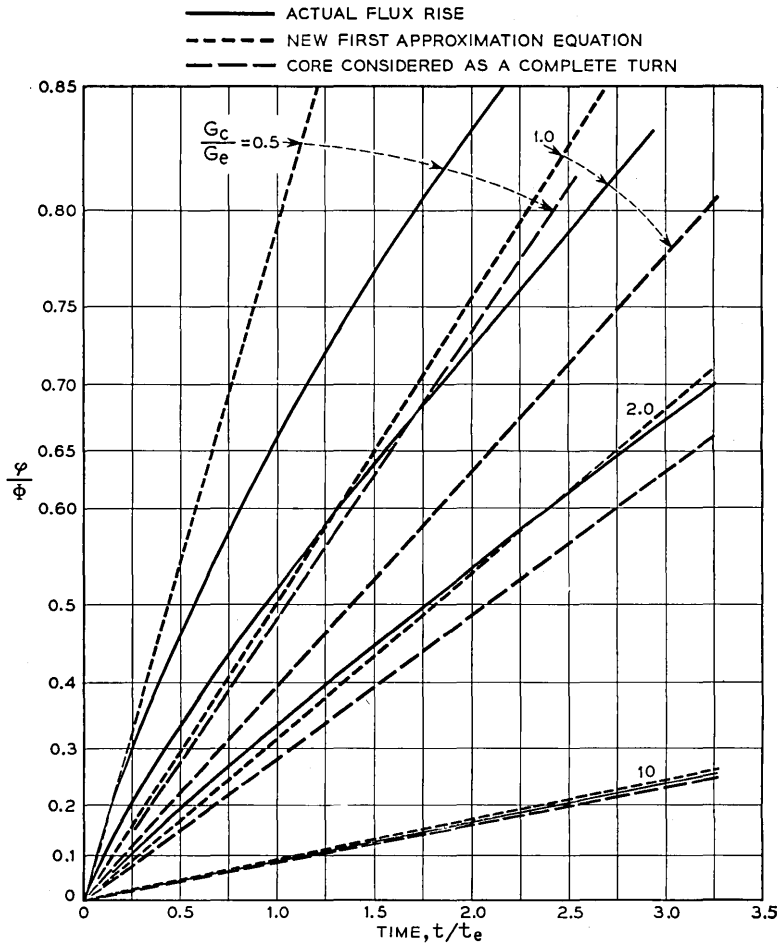


Fig. 25 — Comparisons of equations for flux rise.

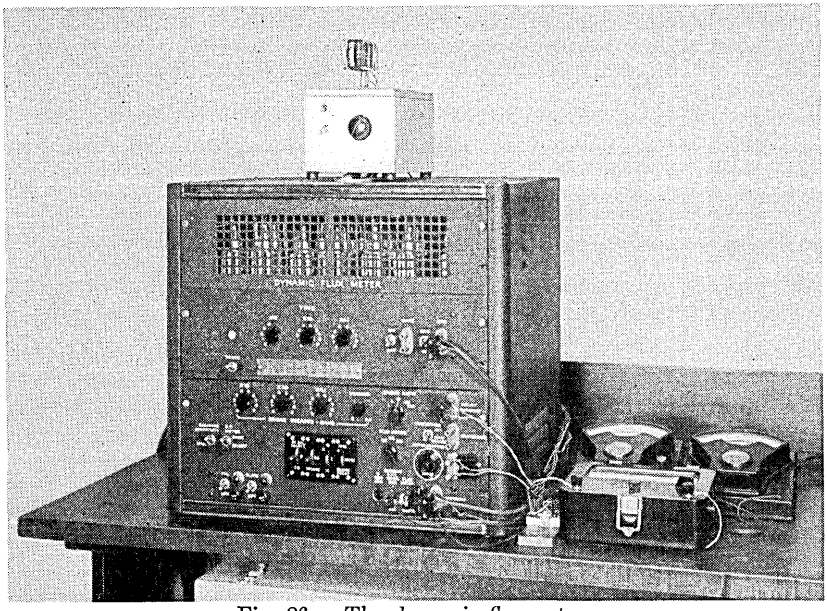


Fig. 26 — The dynamic fluxmeter.

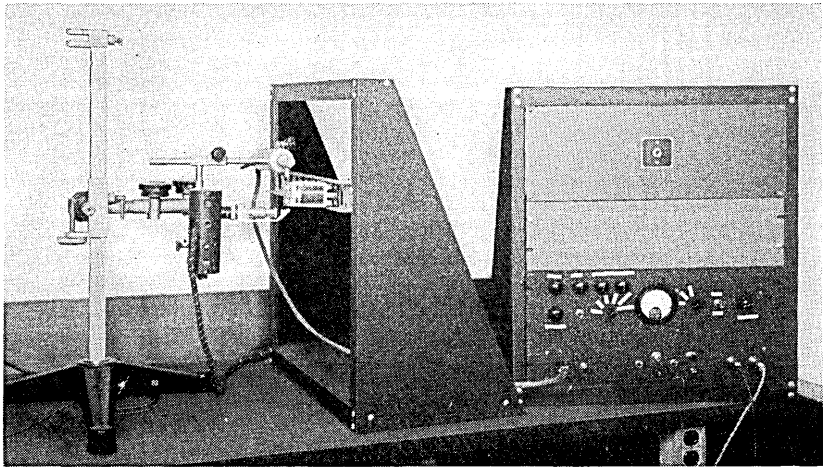


Fig. 27 — The optical probe and associated dc amplifier system.

Thus we have shown the new expression more accurately represents the eddy current effect on the initial flux rise for speed coils, and for most relays the overall accuracy will be within 2 per cent, and all relays should be within 4 per cent, an error reduction of about two-thirds compared to the older approximation.

EQUIPMENT

The dynamic fluxmeter is shown in Fig. 26. Two panels, each with a subchassis attached, comprise the set. The dc instrument and a supplementary voltmeter and ammeter are on the bench beside the cabinet. The upper panel contains the timing system, with the time selection controls. The lower panel includes the power supply and the test relay circuit controls, including the mercury contact relays. The control keys arrange the test relay circuit for any test condition.

The optical probe, a test relay, and the dc amplifier system are shown from left to right in Fig. 27. The photocell is in the shielded container above the relay, with the lamp and lens system below. A right angle prism turns the light beam into the vertical, to pass between the vanes on the relay.

The upper panel of the bench rack is the −250-volt supply. The lower panel includes the +250-volt supply, the three dc amplifiers and the magnetically shielded air-core output transformer. Its secondary is connected to the fluxmeter through a shielded cable. The controls are for zero setting the dc amplifiers, for selecting the amplifier gain and whether a displacement or velocity measurement is to be made.

REFERENCES

1. E. L. Norton, Dynamic Measurements on Electromagnetic Devices, A.I.E.E. Trans., **64,** p. 151, April, 1945.
2. Keister, Ritchie and Washburn, *The Design of Switching Circuits,* D. Van Nostrand, 1951.
3. Otto Schmidt, Thermionic Trigger, J. Scientific Instr., **15,** p. 24, 1938.
4. Richard Weissman, Stable Ten-Light Decade Scaler, Electronics, **22,** May, 1949.
5. M. Artzt, Survey of DC Amplifiers, Electronics, Aug. 1945.
6. Bode, *Network Analysis and Feedback Amplifier Design,* D. Van Nostrand 1945, p. 162.
7. R. M. Bozorth, *Ferromagnetism,* D. Van Nostrand, 1951, Chapter 17.

Selenium Rectifiers—Factors in Their Application

By J. GRAMELS

(Manuscript received July 1, 1953)

Selection of the proper selenium rectifier stacks for best results in the design of dc power supplies involves consideration of characteristics not ordinarily found in published data. This paper describes the data required for the selection of selenium cell sizes and cell combinations, shows typical voltage-current characteristics, and gives the results of extensive life test data necessary for evaluating the life expectancy of the product. The life test data indicate that there are substantial differences in the life expectancy of selenium stacks as manufactured by various companies in this country. Shorter life can be anticipated as the rms cell voltage ratings are increased. In addition, the life is affected by the current density and the temperature at which the selenium cells are operated.

INTRODUCTION

Since their introduction in this country about 1939, selenium rectifier stacks have proved to be a useful means of converting ac power to dc power for Bell System applications. These applications vary from 2 to 2,500 volts and in power sizes from a few watts to 10 kilowatts. Properly designed, the selenium rectifier has a relatively long-life expectancy and requires a minimum amount of maintenance. For these reasons, selenium rectifier stacks are widely used in telephone plants for battery charging, relay operation, plate and filament supply for vacuum-tube amplifiers, bias supplies, telegraph and teletypewriter circuits.

Up to 1952, about 245,000 rectifiers of all types have been manufactured for the Bell System. These include tungar, copper-oxide, vacuum tubes, thyratrons and selenium types. Of this total, about 25 per cent are of the selenium type.

Although Bell Laboratories studies of selenium rectifier stacks date from 1939, rectifiers using such stacks did not enter the telephone plant until 1945. From 1940 to 1945, however, selenium stacks were designed

widely into communication systems for World War II military projects. Since 1945, selenium rectifier power supplies have increased rapidly. For example, in 1945, out of a total of 9,000 rectifiers of all types, about 1,000 or 11 per cent used selenium. In 1951, about 30,000 out of 45,000 rectifiers, or 67 per cent, were of the selenium type.

There are many companies in this country who manufacture selenium stacks but the quality and behavior of the various manufacturers' products show considerable variation, particularly in regard to life expectancy. For many years the Laboratories has carried on an extensive testing program to evaluate properly the various suppliers' rectifiers. As a result of these continuing investigations, the Laboratories is in a position to select cell sizes and combinations of selenium rectifier stacks for use in new power applications. Specifications then are written on one or more suppliers. These specifications have a three fold purpose: (1) They are used as a purchasing and inspection document, (2) they cover the electrical and mechanical requirements necessary for proper electrical design and equipment layout, and (3) they are useful in maintain ing records of the electrical and mechanical characteristics.

CELL MANUFACTURE AND STACK ASSEMBLY

A selenium rectifier cell is an elementary rectifying device having one positive electrode, one negative electrode and one rectifying junction. Cells are made by coating a chemically treated base plate, usually aluminum, with a thin layer of purified selenium to which a halogen element (chlorine, iodine or bromine) has been added. This mixture is applied to the base plate by one of the following methods:

1. Sprinkled or dusted on and subjected to heat and pressure.
2. Deposited by an evaporation process in an evacuated chamber.
3. Dipped in molten selenium and spun.

The selenium then is converted to the desired crystalline structure by heat treatments. During the heat treatment, a blocking or barrier layer is formed on the exposed surface of the selenium. This layer is further developed by various chemical means, which are "trade secrets" with each manufacturer. A thin layer of low melting point alloy, the front electrode, is then sprayed on the selenium.

The manufacturing process is completed by electrically "forming" the cells by applying a pulsating dc voltage in the non-rectifying direction for a specified time interval.

Selenium cells are assembled on an insulated bolt or stud, with individual cells separated by metal spacer washers to allow free passage of air for cooling the assembly. Contact terminals are brought out in various

arrangements for series, parallel or series-parallel connection of the cells as required.

The completed assembly is designated as a rectifier stack. A rectifier stack is a single structure of one or more rectifier cells.

Small size cells (less than 1″) have no center hole for mounting. These cells are assembled without spacer washers in glass, metal, phenol fibre or bakelite tubing. Terminal leads extend outside these enclosures.

SYMBOLIC NOTATION

The combination of cells on a stack is described by a sequence of four symbols written a-b-c-d with the following significances:* (a) number of rectifying elements; (b) number of cells in series in each rectifying element; (c) number of cells in parallel in each rectifying element; and (d) symbol designating circuit or stack connections.

The symbols for the more common types of stack assemblies are shown schematically in Fig. 1.

A *basic* selenium stack is defined as a stack having a single selenium cell in each rectifying element. For instance, a 4-1-1B stack is a basic full-wave single-phase bridge rectifier stack with one selenium cell in each of the four rectifying elements.

The total number of selenium cells on any stack is the product of the three numbers indicated. For example, a single-phase full-wave bridge stack with three cells in series and two cells in parallel per rectifying element would be designated as a 4-3-2B stack assembled with $4 \times 3 \times 2$ or 24 cells.

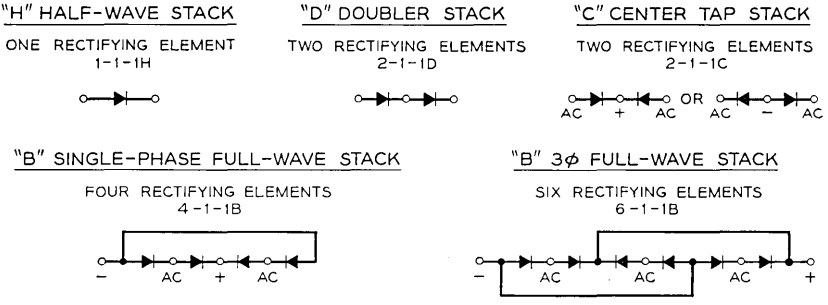


Fig. 1 — Stack assembly symbols. Color coding of terminals (AIEE and NEMA Standard): yellow for ac, red for plus dc output, and black for negative dc output.

* This method of specifying stack assemblies has been standardized by both the National Electric Manufacturer's Association and the American Institute of Electrical Engineers.

DESIGN CONSIDERATIONS

The proper selection of selenium stacks for use in dc power supplies involves a number of important factors that must be carefully considered.

(1) Circuit requirements must be carefully analyzed so that alllowance may be made for the normal variations in the voltage-current characteristics of each manufacturer's product as well as variations that exist for the same stacks processed by different manufacturers. For a fixed ac input voltage, differences in the forward voltage drop may vary the dc output voltage at least $\pm$ 3 per cent from the mean value. If this cannot be tolerated, selenium cells have to be carefully graded and selected to obtain uniformity or other circuit adjustments have to be provided. Special selection of cells, obviously, will increase the cost of the product.

(2) The engineer must take into account the magnitude of the changes in the voltage-current characteristics of the rectifier stacks over the specified temperature range of his project. At very low temperatures, output voltages may be 5 to 10 per cent lower than at normal room temperatures. For high temperature operation, the stacks must be properly derated for both current and voltage to prevent overheating and rapid failure.

(3) Selenium rectifiers age with time. Compensation for this aging should be provided if load requirements warrant. The project engineer should determine what life he expects or requires of the application. For military applications life requirements may vary from minutes to thousands of hours. On other applications, such as telephone and elevator installations, it is desirable to design selenium rectifiers for life expectancies of ten to twenty years or more.

(4) The equipment engineer must anticipate the differences in mechanical details of the same stack assembled by different suppliers. There is no standardization in the selenium industry regarding cell sizes or mechanical details such as the overall length and height of the stack and particularly the type of mounting. However, a committee for the National Electrical Manufacturers' Association is attempting to standardize these mechanical details so that stacks assembled by different suppliers will be mechanically interchangeable.

(5) Unless otherwise specified, rectifying stacks are coated with various types of paints and varnishes for protection against moisture in normal conditions of humidity. For military projects and other applications where selenium rectifiers may be exposed to high humidities, fungus, salt or other corrosive atmospheres, the rectifier stacks must be

provided with a more suitable type of protective coating or finish. These finishes are available from most manufacturers.

(6) When selenium stacks are mounted in cabinets or housing with other heat-generating devices, they should be arranged in such a manner that heat from the other components does not reach the rectifier stacks. The stacks should be mounted below the other components, and in such position that the free flow of air through the rectifier stack is not impeded. The stack should be mounted with the assembly stud in the horizontal, not vertical, position. When two or more stacks are mounted in the same housing, they should be in a horizontal plane with each other, or properly staggered. Cabinets should be provided with louvers to dissipate the heat within the enclosure.

ELECTRICAL RATINGS

The electrical ratings of selenium rectifier stacks are based on their voltage, current and thermal characteristics. All three must be considered carefully for initial design purposes, as any one can affect life expectancy.

Voltage Ratings

The voltage rating usually is expressed as the "reverse voltage rating." It is the maximum rms sinewave voltage above which an excessive reverse current would flow and overheat the cell, causing breakdown. However, the important consideration establishing the rating is the peak voltage applied to the cell. If the applied voltage differs significantly from a sine-wave, it is important that the applied peak voltage shall not exceed 1.41 times the rated rms voltage.

When selenium cells originally were manufactured in this country, their rms reverse voltage ratings ranged from 14 to 18 volts. Ratings later increased to 26 volts. One manufacturer already has successfully produced 33-volt cells for several years; and another supplier announced recently a 40-volt cell. Cells rated at higher voltage have been produced in the laboratory. For non-critical applications, such as low-cost radio and television sets, some selenium manufacturers make cells rated at 45 volts rms. Generally, in such applications, a high reliability product is not required and the units may have a relatively short life.

The nominal dc output voltages obtained from various common *basic* rectifier stacks assembled with cells rated at 18-, 26- and 33-volt rms are listed in Table I. These ratings apply for stacks operating at rated dc output current into a resistance load.

Current Ratings

The current rating of selenium cells is based upon a normal current density of 0.32 ampere for each square inch of active rectifying surface for a single-phase full-wave bridge rectifier stack (4-1-1B) operating into a resistance load. The rating applies to stacks in which the cells are separated by spacer washers so that the cells are cooled by convection air currents. Small size cells mounted without spacer washers are rated at a lower current density.

The dc output current ratings for various selenium cell sizes are listed in Table II.

These ratings are based upon continuous operation in ambient temperatures up to +35°C with unrestricted ventilation for convection cooling of the stacks. For *battery or condenser loads* on single phase circuits, the above current ratings usually are derated to 80 per cent of the above values.

Thermal Ratings

Voltage and current ratings usually are based upon operation of the rectifier stack in a normal ambient temperature of +35°C. It is important to note, however, that for selenium cell application, ambient temperature is defined as the temperature *immediately* surrounding the stack within the equipment enclosure, not the temperature area where equipment is installed. Stacks operating at rated current and voltage into a resistance load have temperature rises ranging from 15 to 30°C when measured with a thermocouple in direct contact with the cell. The temperature rise depends upon the manufacturing techniques used by the various companies as well as the spacing of the selenium cells on the assembled stack. For long life expectancy, the actual cell temperature should not exceed +75°C.

TABLE I — NOMINAL DC OUTPUT VOLTAGES OF RECTIFIER STACKS

	Basic Stack	DC Output Volts		
Cell Rating, Volts rms		18	26	33
Single Phase				
Half-wave	1-1-1H	6	9	12
Full-wave				
Center tap	2-1-1C	6	9	12
Bridge	4-1-1B	12	18	24
Three-Phase				
Full-wave bridge	6-1-1B	19	30	38

For operation in ambient temperatures above +35°C, the voltage or current ratings, or both must be reduced to limit the cell temperature (ambient plus heat rise) to +75°C. The de-rating factors for voltage and current vary somewhat from supplier to supplier, but Fig. 2 shows typical de-rating curves for operation of selenium stacks above a +35°C ambient.

Selenium rectifiers can be operated at cell temperatures above 75°C, but aging accelerates with temperature and excessive temperatures will result in rapid failure.

A special problem today is the application of selenium rectifiers for high temperature military uses. Many military projects are requesting designs to operate at ambient temperatures of +70 to +90°C. Unfortunately very little data about rectifier operation and life expectancy under these conditions have been obtained up to now.

VOLTAGE — CURRENT CHARACTERISTICS

To demonstrate the non-linear characteristics of selenium rectifier cells, Fig. 3 illustrates representative dynamic voltage-current curves showing the open circuit and short circuit characteristics of bridge connected stacks composed of various cell sizes rated at 26 volts, rms. This rating was chosen since all manufacturers presently are producing cells of this type.

Cells with higher voltage ratings generally have a slightly greater forward voltage drop.

TABLE II — DC OUTPUT RATINGS

Cell Size	Area Square Inches*	Continuous DC Amps at +35°C		
		Single-Phase		Three-Phase
		Half wave	Bridge & C.T.	Bridge
¼″ Dia.	0.049	0.006	0.012	0.018
½″ Dia.	0.196	0.025	0.050	0.075
1″ x 1″	0.56	0.090	0.18	0.270
1¼″ x 1¼″	1.1	0.175	0.35	0.525
1½″ x 1½″	1.7	0.270	0.54	0.810
2″ x 2″	3.1	0.500	1.0	1.50
3″ x 3″	7.0	1.10	2.2	3.30
4⅜″ Dia.	12.5	2.00	4.0	6.00
5″ x 5″	21.2	3.40	6.8	10.2
4¼″ x 6″	21.2	3.40	6.8	10.2
5″ x 6″	26.2	4.20	8.4	12.6

* Approximate active rectifying area (varies with suppliers)

The data shown in Fig. 3 and subsequent data are plotted for *basic* 4-1-1B stacks. Tests were made with 60-cycle sinusoidal supply voltage. The "forward voltage drop" is expressed as the rms volts required to produce a specified current in a moving coil dc ammeter connected directly to (short circuiting) the output terminals of the rectifier stack. For stacks with more than one cell per element, the voltage drop is obtained by multiplying the observed drop in Fig. 3 by the number of cells in series per rectifying element.

The reverse current is measured by applying a specified rms voltage to the ac terminals with the dc terminals open circuited and noting the rms input current after the current has stabilized. When selenium stacks have been "off voltage" for some time, a relatively high reverse current is obtained for the first few seconds. The current then decays approximately exponentially. Usually, the current will stabilize after voltage has been applied for 5 to 10 minutes.

It should be emphasized that these curves are only typical characteristics. Selenium cell manufacture requires individual testing of each cell before assembling into a stack. Cells are graded by their electrical characteristics. Large variations exist between the lowest and highest grade. Each manufacturer sets up his own standards regarding the variations

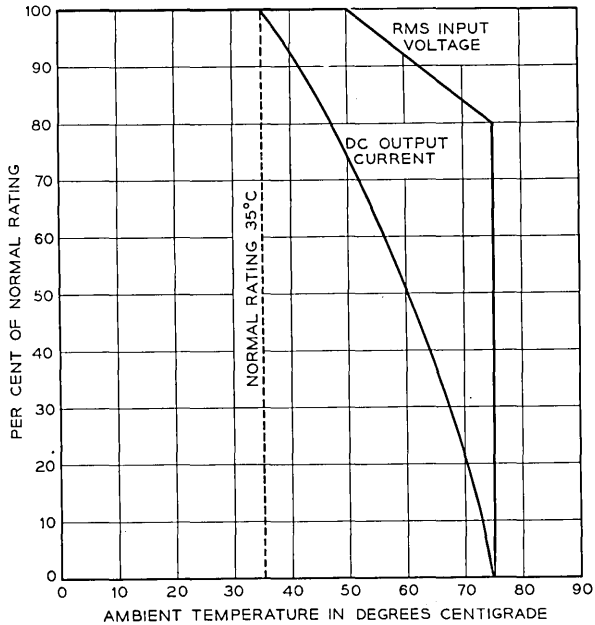


Fig. 2 — Temperature de-rating curves for long-life expectancy.

of characteristics in a given grade. More than one grade of cell may be assembled in a rectifier stack. The art of manufacturing selenium cells is such that, considering production over a yearly period, differences in the forward voltage drop at rated current may vary as much as ±30 to 50 per cent from the mean value. In the reverse direction, a larger per cent spread exists in the reverse current, particularly in cells or stacks made by different suppliers.

Fig. 4 shows dynamic characteristics plotted on a linear scale to illustrate variations of the forward voltage and reverse current characteristics of selenium rectifier stacks that are processed by two different suppliers. Variation of this magnitude exist, not only from supplier to supplier, but also may occur in a particular suppliers' product.

Selenium rectifier stacks in common with other semi-conductor rectifiers, have a negative temperature coefficient of resistance in the forward direction. The forward voltage drop at a specified current decreases as the ambient temperature increases (see Fig. 5). In the reverse direction, the reverse current decreases as the temperature is lowered to approximately −20°C. Below this temperature, there is no apparent change in the current except at the higher voltages. At the higher voltages, the current again tends to increase.

STACK DESIGN

The dc output voltage-current characteristics for various ac input voltages of *basic* rectifier stacks (one cell per rectifying element) are represented in Figs. 6, 7 and 8. The data were obtained by maintaining a constant 60-cycle rms voltage at the stack input terminals and do not take into account transformer regulation or other regulating devices that may be used.

Fig. 6 shows the single phase full-wave bridge characteristics for *resistance* loads.

Fig. 7 shows the single-phase full-wave bridge characteristics for *battery* loads. It will be observed that, with single-phase circuits for a given dc output voltage, lower ac input voltage is required for a battery load than for a resistance load. Capacitor loading is somewhat similar in output characteristics to battery loading, but the value of output voltage is dependent upon the magnitude of the capacity, the quality of the capacitor, and the current drawn by the load. For battery loading, the output voltage is dependent upon the type of battery, the condition of the battery, the battery voltage and the charging current rate required.

For these reasons, it is difficult to accurately predict the exact input-

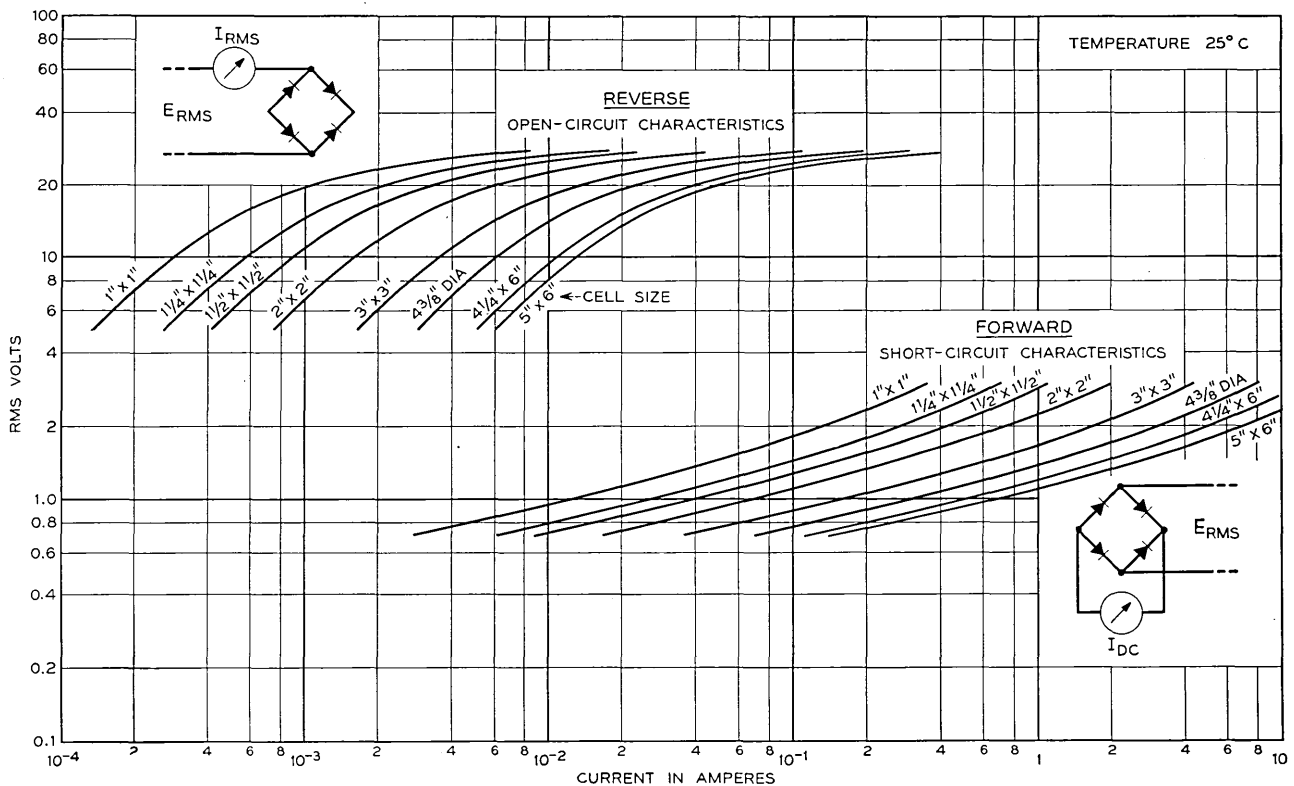


Fig. 3 — Typical dynamic voltage-current characteristics of 4-1-1B stacks with 26-volt cells.

output characteristics of stacks on capacity or battery loads. For precise design information, laboratory tests under the specified conditions should be made to determine the actual slope of these characteristics.

Fig. 8 shows the characteristics for a three-phase full-wave bridge circuit. These characteristics apply to both resistance and battery loads.

To design a selenium rectifier stack, the following procedure may be used:

1. Refer to Table I for the dc output voltage rating for the particular circuit application.

2. For the specified dc output current select the proper cell size from Table II. (If the current exceeds that of a single cell, additional cells may be connected in parallel.)

3. The ac input voltage to the stack for the required dc output voltage and current is then obtained from Figs. 6, 7 or 8.

The following example illustrates this method of stack design.

Example: To design a single-phase full-wave bridge rectifier stack to supply 1.0 ampere dc at 48 volts into a resistance load. From Table I, the highest dc output voltage for a basic stack is 24 volts for cells rated

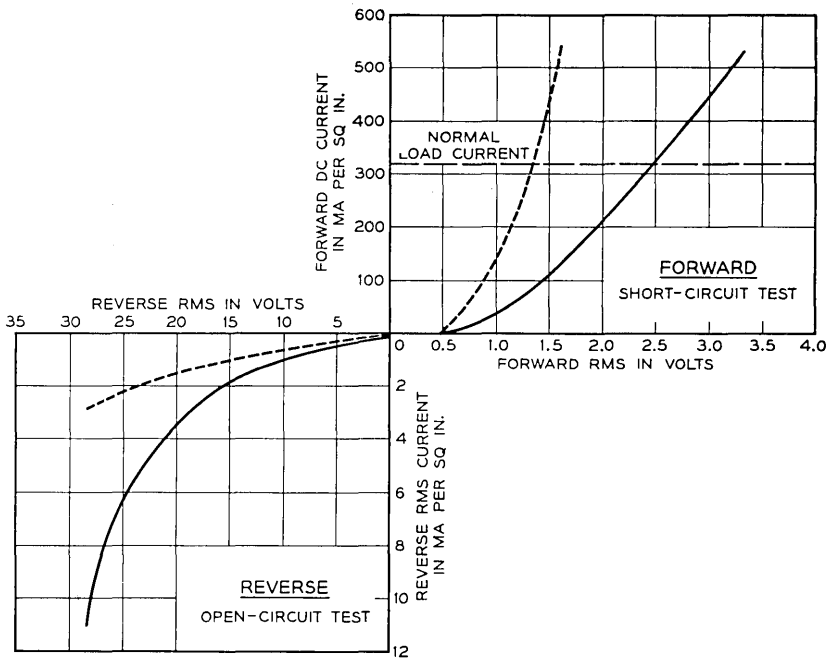


Fig. 4 — Typical dynamic voltage-current characteristics of 4-1-1B stacks with 26-volt cells processed by two different suppliers.

at 33 volts, rms. Therefore, $48/24$ or 2 cells in series in each of the four rectifying elements of the full wave bridge circuit will be required. From Table II, a 2-inch square cell is needed to carry the specified load of 1 ampere dc. As previously described under "Symbolic Notation," the rectifier stack would be designated as 4-2-1B combination using 8 cells.

Selenium cells rated at 18 or 26 volts rms also could be used, but the total number of cells in the stack would be increased as shown in Table III.

Obviously, the higher the cell voltage rating, the less total cells and consequently a smaller and lighter stack is obtained. However, other factors, particularly life expectancy, must be evaluated as described later in the text.

After the stack cell combination has been established, the required ac input voltage is obtained from Fig. 6. At 100 per cent normal load current (1 ampere dc from Table II) and 24 volts dc, the ac voltage for a *basic* stack is about 30 volts, rms. Since there are two cells in series in each rectifying element, 2 × 30 or 60 volts rms input voltage would be required for a new stack.

If cells rated at 26 volts rms are selected, three cells in series would be required. For the specified 48 volts dc output, $48/3$ or 16 volts dc per series cell, is obtained. Again referring to Fig. 6, an input voltage of 20.5 volts per cell or 61.5 volts total is required. This compares to 60 volts for the previous illustration. The higher input voltage results from the added voltage drop of the additional cells in the rectifier stack.

This stack design is satisfactory for operation in ambient temperatures not exceeding +35°C. For operation at higher ambients, the stack should be de-rated as shown in Fig. 2. For example, if the stack is required to operate in an ambient of +60°C, the dc output current rating would be 50 per cent of the normal rating. The 2″ square cell previously

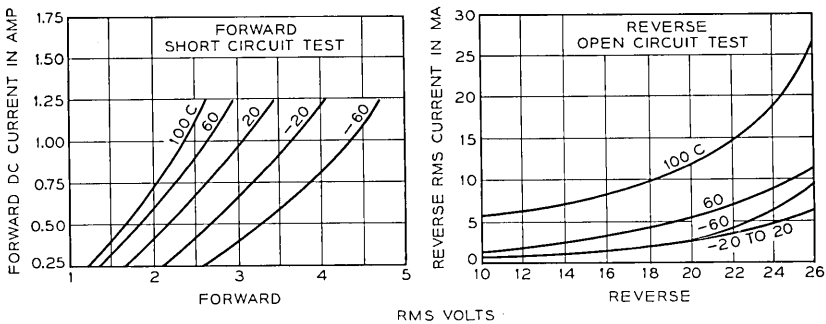
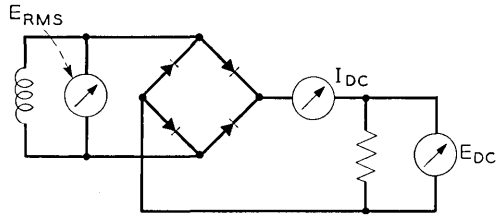


Fig. 5 — Temperature characteristics of 4-1-1B stacks with 2″ x 2″ cells rated at 26 volts rms and 1.0 ampere dc.



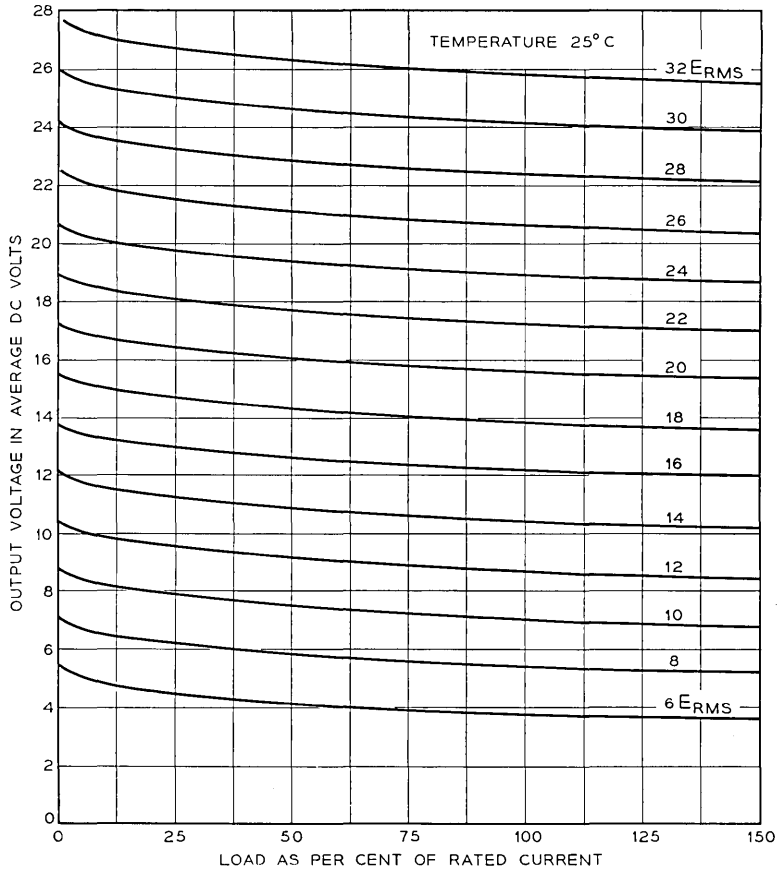
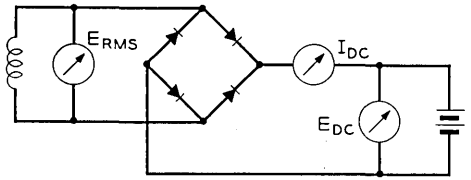


Fig. 6 — Typical output voltage-current characteristics. 60-cycle single-phase full-wave bridge circuit. Basic stack 4-1-1B. Resistance load.



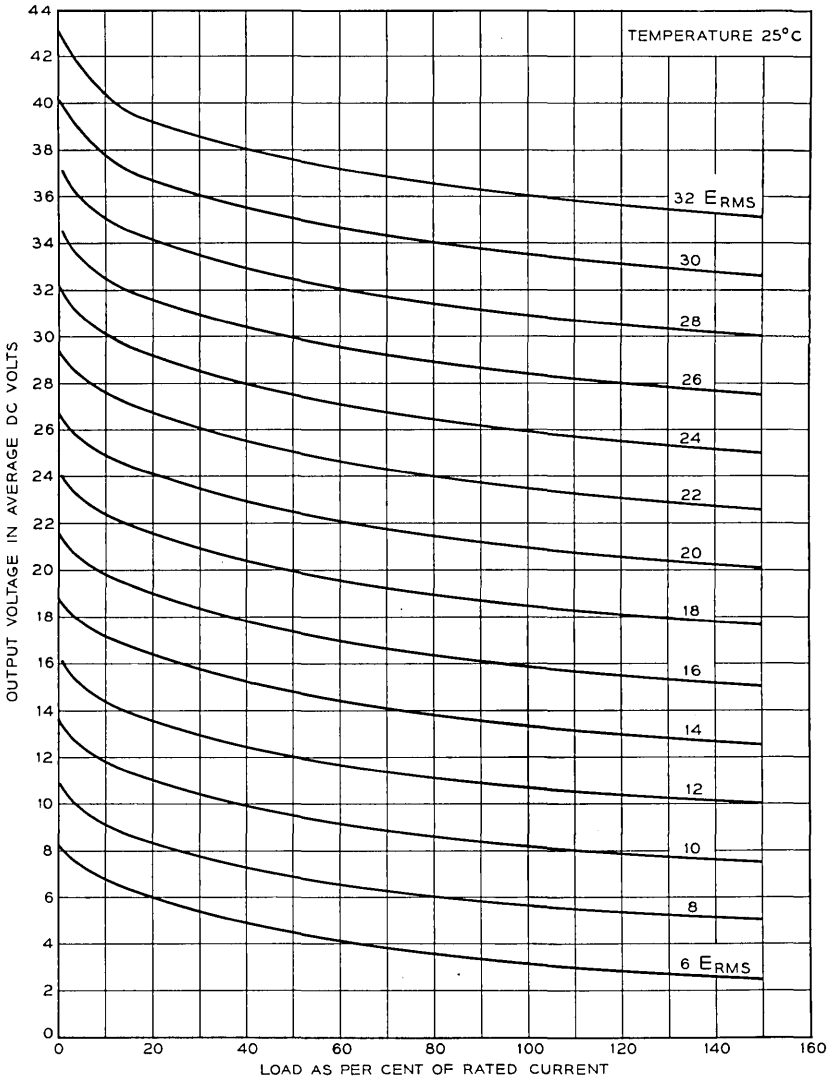
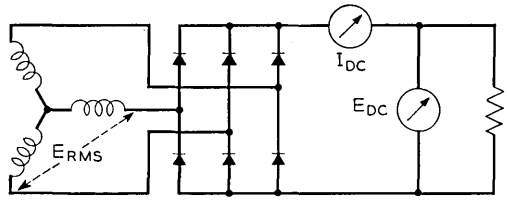


Fig. 7 — Typical output voltage-current characteristics. 60-cycle single-phase full-wave bridge circuit. Basic stack 4-1-1B. Battery load.



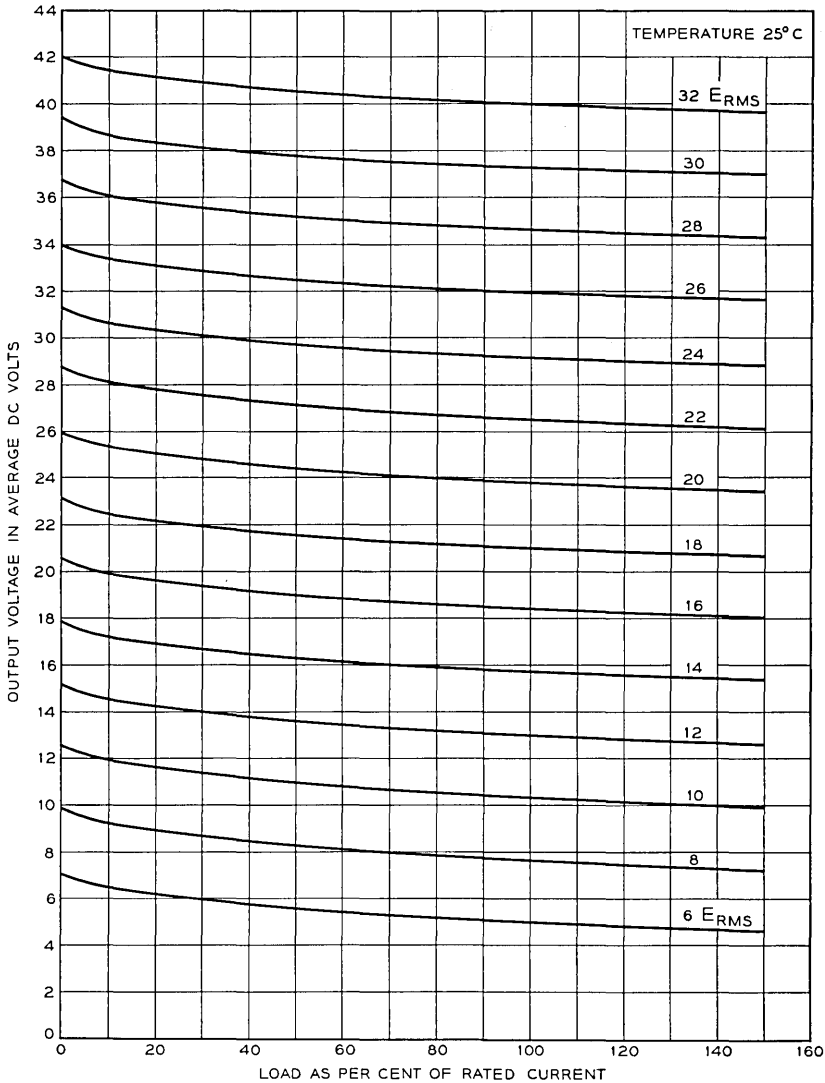


Fig. 8 — Typical output voltage-current characteristics. 60-cycle three-phase full-wave circuit. Basic stack 6-1-1B. Resistance and battery loads.

selected would now be rated at 0.50 × 1 or 0.5 ampere. Therefore, for the specified 1.0 ampere load, two cells in parallel are required or the next larger cell size, (3″ square) could be selected. The 3″ square cell would be rated at 0.50 × 2.2 or 1.1 ampere.

At 60°C ambient, the input voltage for new or aged stacks should not exceed 92 per cent of the normal rms voltage rating.

An alternative method of calculating the input voltage required for any stack cell combination may be obtained by the use of the formulas shown in Table IV.

Since the selenium rectifier stack is a non-linear device whose characteristics vary depending upon the circuit and operating conditions, the values given in Table IV are not precise but they are reasonably accurate for most design purposes.

To illustrate the formulas method, let us determine the input voltage for the previous example, that is, a single-phase full-wave bridge stack to supply 1.0 ampere dc at 48 volts into a resistance load. As previously determined, the stack cell combination would be a 4-2-1B circuit. The

TABLE III — STACK DESIGN

Cell Rating Volts rms.	Cell Combination	Total Cells
33	4-2-1B	8
26	4-3-1B	12
18	4-4-1B	16

TABLE IV — CALCULATING THE INPUT VOLTAGE REQUIRED FOR ANY STACK CELL COMBINATION

Rectifier Circuit	Formulae	K	
		Resistance Load	Battery or Capacity Load
Single-phase Half-Wave	Eac = KEdc + nDV	2.3	1.0
Single-phase Full-Wave, Center TAP*	Eac = KEdc + nDV	1.15	0.80
Single-phase Full-Wave Bridge	Eac = KEdc + 2nDV	1.15	0.80
Three-phase Full-Wave Bridge	Eac = KEdc + 2nDV	0.74	0.74

Eac = Stack input volts, rms.
Edc = Average dc output volts.
K = Circuit form factor.
n = Number of cells in series in each rectifying element.
DV = RMS forward voltage drop at specified dc output current (see Fig. 9).
* For center tap circuits, Eac is the voltage to the mid-tap on the transformer.

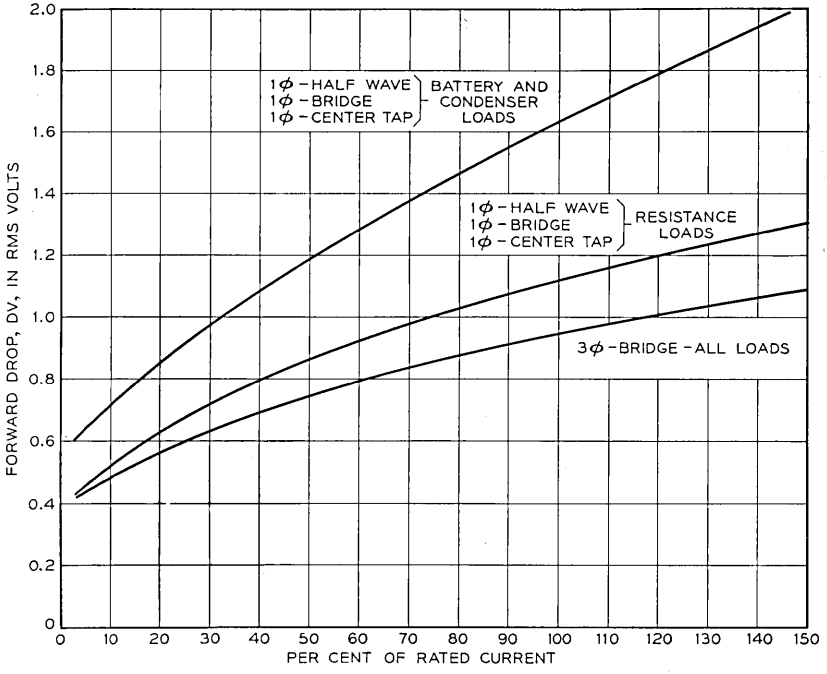


Fig. 9 — Dynamic forward voltage drop per cell.

forward rms voltage drop per cell (DV) obtained from Fig. 9 is 1.10 volts at 100 per cent normal rated current. From the formulas in Table III, the input voltage is:

$$Eac = 1.15Edc + 2nDV,$$
$$Eac = 1.15 \times 48 + 2 \times 2 \times 1.1 = 59.6 \text{ volts rms.}$$

This input voltage is required for a new or unaged stack. Since the stack ages under service operating conditions, additional input voltage will be required to maintain the original dc output. It is considered good practice to design for a 100 per cent increase in the initial forward voltage drop (DV) of the stack. The aged stack, then would require

$$Eac = 1.15Edc + 2(2nDV),$$
$$Eac = 1.15 \times 48 + 2(2 \times 2 \times 1.1) = 63 \text{ volts rms.}$$

The amount by which an increase of 100 per cent in the forward voltage drop or forward resistance will change the output is dependent upon the design of the circuit. It depends upon the ratio in per cent of the forward rectifier stack resistance to the total circuit resistance including

the transformer, ballast, load, etc. For most practical design considerations, one or more aging taps on the input transformer to provide 5 to 10 per cent additional voltage, will compensate for the forward aging.

It should be emphasized that the curves in Figs. 6, 7, 8 and 9 are based upon empirical data obtained on new rectifier stacks and the characteristics will vary slightly depending upon normal manufacturing variations in the voltage current characteristics mentioned earlier in this article.

AGING

Selenium rectifier stacks are subject to aging. Aging is defined as any persistent change (except failure) which takes place for any reason in either the forward or reverse resistance characteristics. The important factor in selenium rectifier aging is the increase in the forward voltage drop which results in a decreased dc output. For normal rectifier applications aging of the reverse current is not critical.

For design purposes, a selenium stack is considered to have reached the end of its useful life when the stack input voltage required to maintain rated output voltage exceeds the rms reverse voltage rating assigned by the manufacturer. Operation beyond this limit will result in overheating of the selenium cells and rapid failure of the stack.

The extent and rate of aging of selenium stacks cannot be predicted mathematically. Aging characteristics must be determined by actual test involving lengthy time consuming projects. To determine whether a given rectifier stack will give satisfactory performance for five years, as an example, tests would have to be conducted for this period. Aging can be accelerated by operation at high temperatures or at load currents above normal, but no accepted correlation exists between this type of aging and that obtainable under normal operating conditions.

Aging data were obtained on sample rectifier stacks obtained from different manufacturers in this country. The samples were set up on life test racks as single phase full wave rectifiers and were operated continuously at normal room ambient temperatures. For the duration of the tests, the 60-cycle ac input voltage to the stacks was kept constant at approximately 10 per cent below the maximum rms voltage rating. The stacks operated into a resistance load. The resistance was selected so that the rectifiers operated at rated dc load currents. The resistance was not changed thereafter.

Long term forward aging characteristics of selenium rectifier stacks

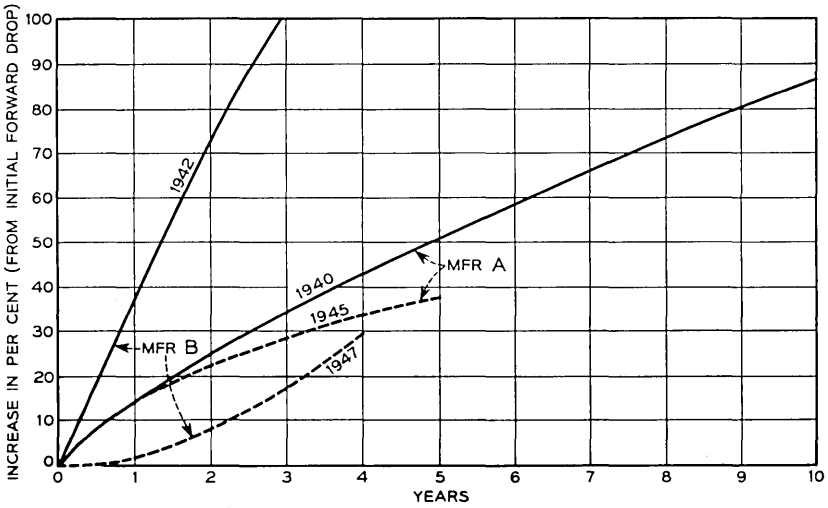


Fig. 10 — Forward aging characteristics of 18-volt cells.

assembled with cells rated at 18 volts rms are shown in Fig. 10.* After 10 years (approximately 90,000 hours) continuous operation, the forward drop on one supplier's product increased approximately 90 per cent. A second supplier's product increased 100 per cent in 3 years under similar operating conditions. Further development programs by each manufacturer resulted in improved aging qualities as indicated by tests made on stacks obtained at a later date.

By introducing changes in their manufacturing techniques, the industry eventually increased the voltage ratings to 26 volts rms per cell. This change in processing, however, raises the question — how has the operating life been affected? Fig. 11* clearly illustrates that although many manufacturers are commercially producing 26-volt cells, the life expectancy is vastly different.

In order to save space, weight and critical materials, manufacturers are attempting to produce cells with still higher voltage ratings. Most companies have had a 33-volt cell development program under way for some time, but again the problem of aging must be considered. Fig. 12* compares the forward aging of 26- and 33-volt cells after 5,000 hours operation. It is evident that in all cases except one, the 33-volt cells age substantially faster than the 26-volt cells. The exception is manufacturer "E" who has not produced 26-volt cells but has commercially processed

* The designations A, B, C, etc. shown on all aging curves are not to be interpreted to represent the same manufacturer's product on each of the figures.

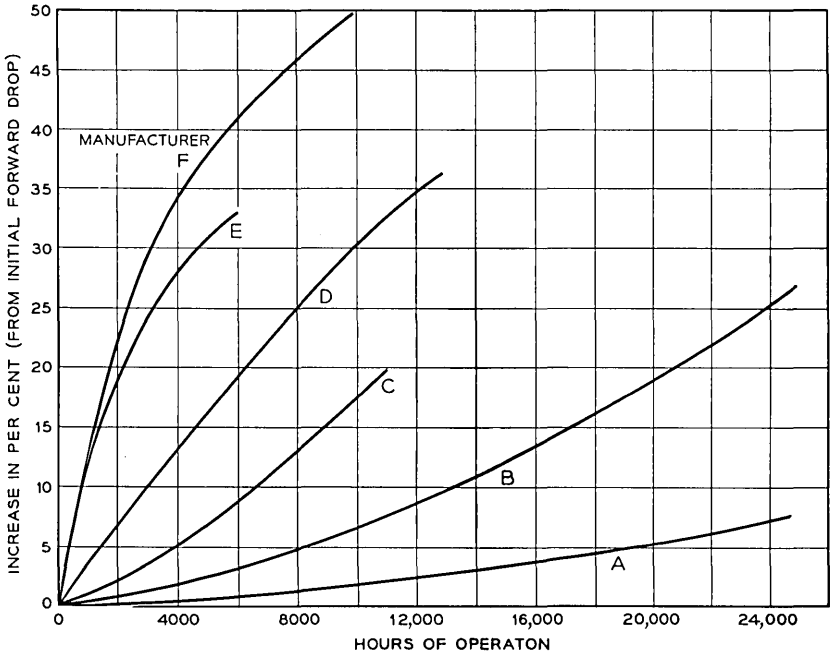


Fig. 11 — Forward aging characteristics of various suppliers 26-volt cells.

33-volt cells for several years. Again, there is a wide difference in the rate of aging in the various manufacturers' products.

Continued aging tests on stacks made by manufacturer "E" show that after 21,000 hours of continuous operation, the forward voltage drop has increased only 8 per cent. Extrapolation of these data indicate a life expectancy of thirty years.

It should be emphasized that the data on the 33-volt cells, with the one exception, were taken on experimental stacks obtained from various companies while they still were in the research and development stages of processing the 33-volt cell. Undoubtedly, further development programs on this type of cell will result in improved aging qualities.

Fig. 13 shows clearly what happens to the aging rate as manufacturers attempt to increase still further the voltage rating. These data are based upon 26-, 33- and 40-volt cells made by one supplier who is commercially producing cells with these ratings. The aging rate is accelerated greatly as the voltage rating is increased.

All the foregoing aging data were taken on stacks operated at rated load currents. A longer operating life may be expected if selenium stacks are selected to operate at lower load currents. As shown in Fig. 14, the aging

at 50 per cent normal current rating is about one half that at rated current. Above normal rating, the aging is accelerated to a greater degree. These data were obtained on a particular supplier's stacks with 26-volt cells, but similar behavior has been observed on other suppliers' products, including stacks assembled with 33- and 40-volt cells.

The behavior of the reverse current during these aging studies indicates that the reverse current generally decreases slightly during the first few months of operation and then remains substantially constant thereafter. A few manufacturers' rectifiers, however, show the opposite effect. The current rises slightly before leveling off. For normal dc power applications, these changes are insignificant.

As previously discussed, selenium stacks are rated for operation at +35°C ambient temperatures and should be derated at higher ambients for normal life expectancy. However, the problem frequently arises — what life can be expected at high ambient temperatures and why cannot aging qualities be predicted over a short time interval under accelerated conditions, such as high ambient temperatures? Fig. 15 shows data obtained on four different suppliers' product when operated at a +80°C ambient temperature. The temperature of the selenium cells under these conditions (ambient plus temperature rise) is about 100°C. For comparison purposes, data are included for similar stacks aged at normal

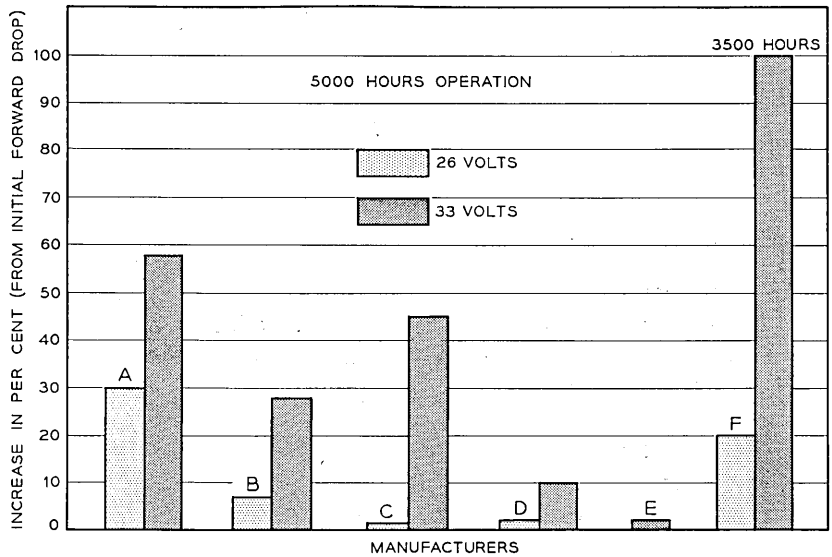


Fig. 12 — Forward aging characteristics. Comparison of 26- and 33-volt cell ratings.

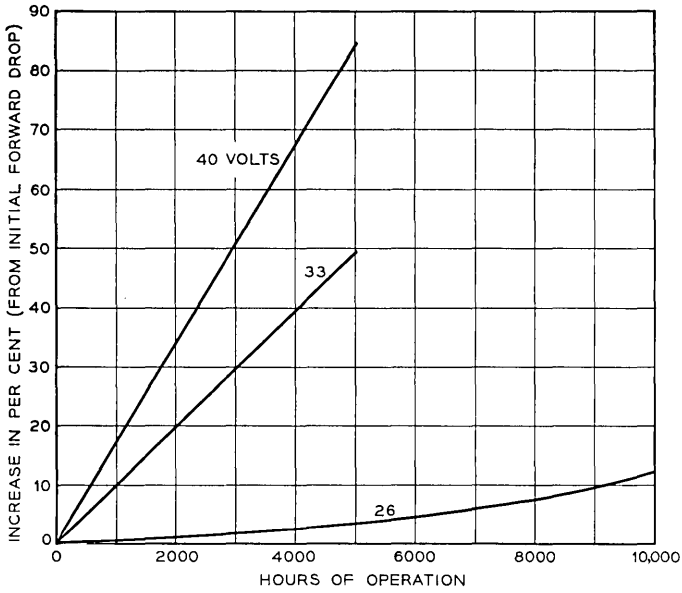


Fig. 13 — Forward aging versus cell voltage rating.

room temperatures. It is readily evident that at +80°C ambient (1) the forward aging rate is increased (2) the different manufacturers' stacks age at different rates and (3) there is no definite correlation of the aging rate with normal operation at room temperature. As previously mentioned under room temperature operation, longer life can be expected if stacks are operated at reduced current loads. At 80°C ambient however, three of the four manufacturers' stacks aged as much at half load as at full load.

CONCLUSIONS

The development and use of selenium rectifier stacks in this country has been quite rapid in the last ten years. The voltage ratings per cell have increased from 14 volts rms to 26 and 33 volts on a commercial basis. The trend in the selenium industry seems to indicate that cell voltage ratings will be further increased.

The proper selection and procurement of selenium rectifier stacks and their application to dc power supplies for large scale industrial or military projects usually require considerably more information than can be obtained from published data in manufacturers' catalogs. Considerable differences have been observed in the performance of similar stacks

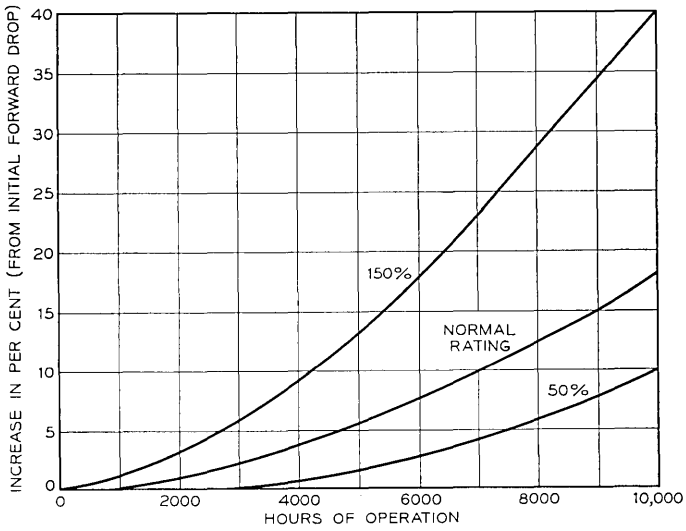


Fig. 14 — Effect of load current on forward aging.

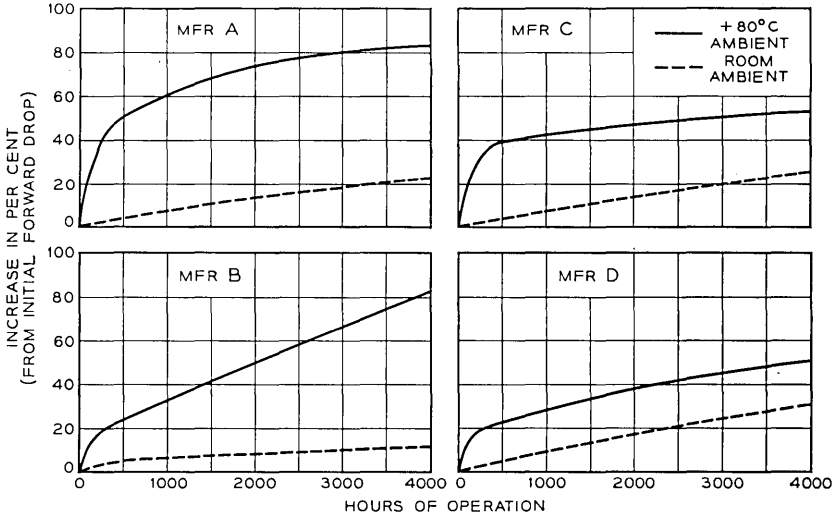


Fig. 15 — Forward aging characteristics. Comparison of room and +80°C ambient temperature operation. 26-volt cells, resistance load, rated dc amps, input voltage — 80 per cent normal for +80°C ambient and 90 per cent normal for room ambient.

furnished by different manufacturers. To appraise properly the qualities of different manufacturers' stacks, the *stability of the life characteristics* should be considered more important than the initial characteristics.

There is very little information published by the selenium manufacturers regarding the life expectancy of their product. Aging appears to be directly related to the individual manufacturing techniques used by each supplier. The life of selenium rectifier stacks seem to decrease as the cell voltage rating is increased. Longer life can be expected if stacks are operated at load currents below the present ratings given in the manufacturers' literature.

Selenium rectifier stacks properly designed and conservatively rated can be expected to give satisfactory performance for 10 to 20 years. Careful consideration of the rate and extent of aging must be evaluated so that proper allowance may be made in the circuit design to obtain maximum life expectancy.

REFERENCES

E. A. Harty, Characteristics and Applications of Selenium Rectifier Cells, Elec. Eng., Oct., 1943.
J. H. Hall, Transformer Calculations for Selenium Rectifier Applications, Elec. Eng., Feb., 1946.
Glen Ramsey, The Selenium Rectifier, Elec. Eng., Dec., 1944.
J. Gramels, Problems to Consider in Approving Selenium Rectifiers, A.I.E.E. Technical Paper No. 53-215. Also published in Communications and Electronics, Sept., 1953.
W. F. Bonner, Advanced Developments in Metallic Rectifiers, Electrical Manufacturing, Oct., 1951.
I. T. Cataldo, Development of 40-volt Selenium Rectifier Plates, Electrical Manufacturing, May, 1952.
S. Niciejenski, Selenium Rectifiers, Radio and Television News, Oct., 1952.
H. K. Henisch, Metal Rectifiers, Oxford, Clarendon Press, 1949.
E. A. Richards, The Characteristics and Applications of the Selenium Rectifier, J. Inst. Elec. Engrs. (London), **88,** Part III, Dec., 1941.

Arcing of Electrical Contacts in Telephone Switching Circuits

Part II—Characteristics of the Short Arc

By M. M. ATALLA

(Manuscript received May 27, 1953)

Results are presented of an experimental study of the characteristics of the short arc in air which is the major cause of contact erosion in telephone switching circuits. Measurements were made of the arc initiation voltage, the voltage drop across the arc and the minimum arcing current. The following are the main conclusions: (1) For "normal" contacts in air, the arc is initiated at a constant field strength of a few million volts/cm up to separations of about 2–3 mean free paths of an electron in air. At larger separations the arc is initiated at the well known spark breakdown potentials of air. In vacuum the linear relation holds for larger separations followed by a transition into a square root relation $V_{ai} = K(d)^{1/2}$. *(2) For "clean" contacts in air, no constant field strength line is obtained for separations as low as 1600Å. Instead, the arc is initiated at the spark breakdown potentials of air, possibly due to adsorbed air molecules or due to breakdown along a longer path at the Paschen's minimum potential. In vacuum, it is speculated that the above square root relation will hold. (3) For "activated" contacts and small separations the arc is initiated at a constant field strength of about* 0.6×10^6 *volts/cm. (4) For "normal" contacts the minimum arcing current increases with an increase in the maximum current during the arc due to surface contaminations and the arc cleaning action. (5) For arc currents above 1.5 ampere and energies of the order of thousands of ergs the cathode determines the arc characteristics.*

INTRODUCTION

The electrical erosion of contacts presents an important problem in the design of telephone switching apparatus. There are several physical phenomena that occur between contacts and contribute to their erosion. The short arc,* which may occur on both make and break of a contact,

* The short arc is characterized by its constant voltage, independent of the current, which is of the order of the ionizing potential of the contact material.

is generally considered to be the major contributor. For illustration, a palladium contact, 10^{-4} cm^3 in volume, will last for more than 10^9 operations* only if the arc energy per operation is less than 2.5 ergs. This is based on an erosion rate of 4×10^{-14} cm^3 per erg.[1] Furthermore, a short arc with a half ampere current, lasting for only one microsecond, will dissipate as much energy as 70 ergs. Contact erosion may also take place, though at much lower rates for the usual ranges of current and voltage in switching circuits, due to molten bridges[2] on contact break and due to glow discharge.[3]

In Part I of this series,[4] was discussed the mechanism of the initiation of the short arc as determined by contact and circuit conditions. Three characteristics of the arc were used in the presentation without elaboration as to their nature: (1) the arc initiation voltage, (2) the voltage drop across the arc, and (3) the arc initiation and the arc termination currents. These characteristics have been the subject of a recent study to which this part of the series is mainly devoted.

In the course of this study, it was found that there should be some repetition of previous work to isolate effects of certain pertinent parameters that were not previously given due consideration.

No attempt is made here to give a complete survey of the related studies in the literature. Only a few publications are referred to as typical references to the subjects discussed.

NOTATION

C Capacitance
E_a Energy dissipated in the arc
F Gross field strength between the contacts: $\frac{V}{d}$

I Current
I_i Arc initiation current
I_{max} Maximum current in the arc
I_m Minimum arcing current or arc termination current

* This is the actual life requirement of some contacts in existing switching circuits.

[1] L. H. Germer and F. E. Haworth, Erosion of Electrical Contacts on Make, J. App. Phys. **20**, p. 1085, 1949.

[2] See for example: J. J. Lander and L. H. Germer, The Bridge Erosion of Electrical Contacts, J. App. Phys. **19**, p. 910, 1948.

[3] F. E. Haworth, Electrode Reactions in Glow Discharge, J. App. Phys. **22**, p. 606, 1951.

[4] M. M. Atalla, "Arching of Electrical Contacts in Telephone Switching Circuits. Part I—Theory of the Initiation of the Short Arc," B.S.T.J., **32**, pp. 1231–1244, Sept., 1953.

K Constant in the relation $V_{ai} = K(d)^{1/2}$
R Resistance
V Voltage
V_{ai} Arc initiation voltage
d Minimum separation between contacts
h Height of a metal bridge formed during one arc
t Time
t_a Arc duration
v Constant voltage drop across the short arc

ARC INITIATION VOLTAGE

Consider the simple contact circuit in Fig. 1 comprising a pair of contacts in series with a resistor R and a variable voltage power supply. By fixing the separation between the contacts and gradually increasing the

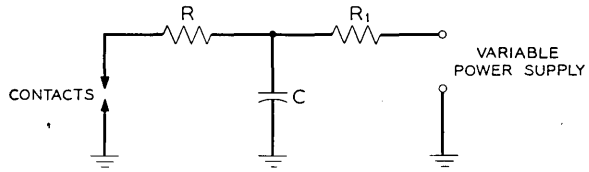


Fig. 1 — Contact Circuit.

voltage, an arc is usually initiated when the voltage reaches a certain value "V_{ai}" called the arc initiation voltage. In general V_{ai} is a function of: (1) separation between the contacts, (2) geometry of the contact surfaces, (3) the surrounding atmosphere, and (4) contact material and its surface.

Our experiments were limited to contacts in atmospheric air with special emphasis on separations of the order of and less than the mean free path of an electron in air. For larger separations the arc initiation voltage follows the well known curve of the sparking potential of air.[5] For the smaller separations where the presence of air molecules would not be expected to affect the arc initiation voltage, it was previously reported that breakdowns occurred at some constant field between 0.6 $\times$ 10^6 and 16 $\times$ 10^6 volts/cm.[1, 6, 7]

In our experiments a cantilever bar, described by Pearson,[6] was used. The setting for zero separation was determined by a 0.1 volt source, a

[5] See for instance J. D. Cobine, Gaseous Conductors, McGraw-Hill, N. Y., p. 162, 1941.
[6] G. L. Pearson, Phys. Rev. **56**, p. 471, 1939.
[7] L. H. Germer and J. L. Smith, J. App. Phys. **23**, p. 553, 1952.

10,000 ohms resistor and a cathode relay oscilloscope. The zero setting could be repeated with a precision of ± 500Å. All the reported results were obtained by fixing the contact separation, raising the voltage and observing the breakdown on a cathode ray oscilloscope. Contacts tested were given one of three different surface treatments:

(1) The contact surface was polished with fine emery paper, washed with methyl alcohol, then exposed to the laboratory atmosphere for a few hours. After about 5 arc discharges, readings of arc initiation voltage seemed to vary at random. Contacts thus treated are referred to as "normal" contacts.

(2) Contacts were subjected to arcing for about 5 minutes at the rate of 15 arcs per second. Arcing was produced by the discharge of a half microfarad condenser at 500 volts through a 10-ohm resistor. Measurements of the arc initiation voltage followed *immediately after* this treatment. These contacts are referred to as "clean" contacts. Their behaviour usually changed to that of "normal" contacts after a short exposure to the laboratory atmosphere.

(3) Contacts were subjected to arcing at the rate of 3 arcs per second for about one hour in air saturated with d-limonene. The arc was produced by discharging a 0.1-microfarad condenser at 50 volts through a 100-ohm resistor. These contacts are referred to as "activated" contacts.[8]

Fig. 2 shows the results obtained with "normal" palladium contacts. Each point represents the average of five readings. The maximum spread was 40 per cent of the average. For separations less than 10,000Å, about two mean free paths of an electron in air at normal conditions, a constant gross field strength line of 3×10^6 volts/cm was obtained. At larger separations the measured arc initiation voltages were essentially the well known sparking potentials of air. Fig. 3 shows the corresponding results obtained with "normal" carbon contacts in air. Below a separation of 15,000Å, the arc was initiated at a constant gross field strength of 2.4×10^6 volts/cm. The maximum spread of the individual points was only 15 per cent of the average.

In the absence of air, it is expected that the constant field strength lines will hold for higher separations.* In Table I, Column 2 are given the measured values of the gross field strengths at which the arc was initiated for a group of "normal" contact materials.

* Recent unpublished measurements by Dr. P. Kisliuk on similar contacts in vacuum have indicated that the constant field strength relation $V_{ai} = F(d)$ holds initially for larger separations and is followed by a gradual transition into a square root relation $V_{ai} = K(d)^{1/2}$ as proposed by Cranberg.[9]

[8] L. H. Germer, Arching at Electrical Contacts on Closure—Part I, J. Appl. Phys. **22**, p. 955, 1951.

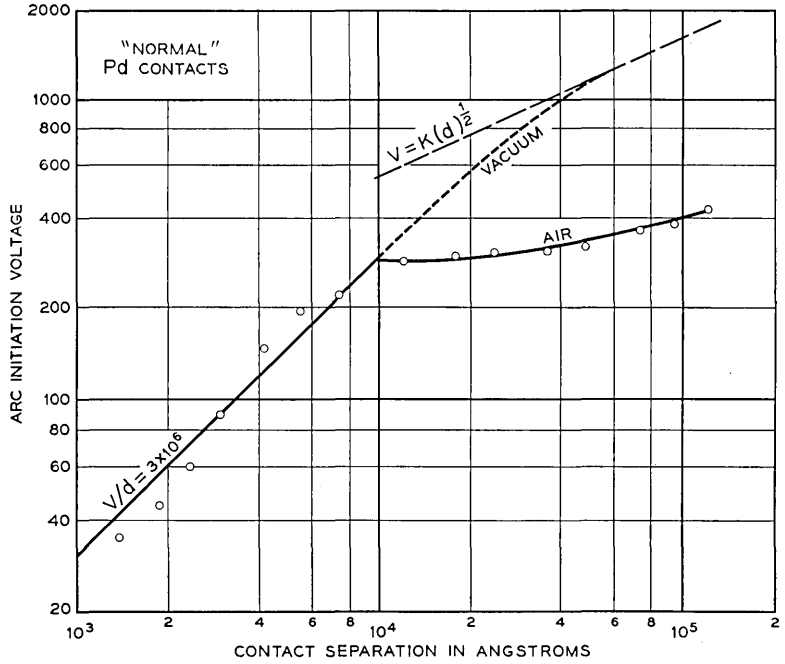


Fig. 2 — Arc initiation voltages for "Normal" palladium contacts.

Due to the spread of the above measurements and their observed dependence on surface exposure and treatment it was suspected that the constant field strength characteristic obtained was due to surface contamination. This was verified by testing contacts cleaned by the heavy arcing process explained above. The results are shown in Fig. 4. The familiar constant field strength line *was not* obtained for separations as low as 1500Å. Instead, the arc was initiated at voltages comparable to the sparking potentials of air. Since the separations were too small, the smallest being three times less than the mean free path of an electron in air, it was thought that the effect was due to some mechanism involving the adsorbed air molecules or due to breakdown along a longer path at the Paschen's minimum potential.[5] In the absence of air, it is, therefore, expected that higher voltages and higher field strengths in excess of 20×10^6 volts/cm, as obtained at 1500Å, will be needed to initiate the arc. It is possible that Cranberg's relationk[9] $V = K(d)^{1/2}$, will hold for separations as low as a few thousand angstroms. In Fig. 4, this relation, with $K = 10^5$ volt. $cm^{-1/2}$, is plotted.

[9] L. Cranberg, The Initiation of Electrical Breakdowns in Vacuum, J. Appl. Phys. **23**, p. 518, 1952.

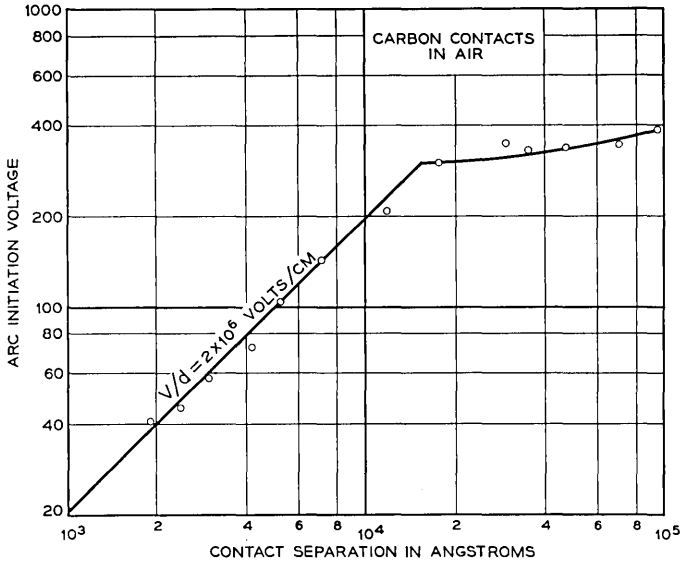


Fig. 3 — Arc initiation voltages for carbon contacts.

For contacts activated in organic vapors, constant field strength lines were obtained. In Fig. 5 results are shown for palladium contacts activated by d-limonene. The average field strength for arc initiation is only 0.6×10^6 volts/cm with a spread of as much as 100 per cent of the average. At separations greater than 50,000Å the arc was initiated by the familiar spark breakdown of air. In vacuum the constant field strength line should hold for larger separations possibly until it intersects Cranberg's line.[9]

Breakdowns at low fields were also observed for metals with inorganic films. For instance Gleichauf[10] obtained a constant field strength line of 0.24×10^6 volts/cm for copper electrodes in vacuum at separations of the order of millimeters. Our measurements on copper contacts at separations less than 10,000Å have shown breakdowns at fields as low as 0.7×10^6 volts/cm. It is concluded that the presence of organic or inorganic films on a contact usually leads to a reduced gross field strength at which the breakdown will occur. The reduction can be by as much as two orders of magnitude. It is possible that this reduction was only apparent and the electrons actually came from the underlying metal by extraction in an intense field set up by the positive ions lying on the surface of the film.[11]

[10] P. H. Gleichauf, Electrical Breakdown Over Insulators in High Vacuum, J. Appl. Phys. **22**, p. 766, 1951.
[11] F. L. Jones, Electrical Discharges in Gases, Nature, **170**, p. 601, 1952.

TABLE I. — ARCING CHARACTERISTICS OF CONTACT MATERIALS

(1) Contact Material	(2) Field strength to initiate the arc for normal contacts. 10^6 Volts/cm.	(3) Short arc Voltage	(4) Minimum arcing current for "clean"* contacts. Amps.
Carbon	2.4	20–43	0.03
Nickel	4.2	12–13	0.4
Palladium	3.0	14–15	1.1
Silver	2.0	11–13	0.8
Tungsten	4.9	12–13	0.7

* For "normal" contacts, the minimum arcing current is less by 50 per cent or more.

The results of the above section are summarized in Fig. 6. The solid lines were actually measured for palladium contacts under different surface conditions. The broken lines are only speculative. For a certain separation and surface condition the arc will be initiated at a voltage as given by the lowest corresponding line in the figure.

VOLTAGE DROP ACROSS A SHORT ARC

The short arc may be defined as a discharge of electricity between electrodes with a voltage drop of the order of the minimum ionizing potential of the atoms of the electrodes.* Furthermore, due to the small separation between the contacts and the local high pressure metal vapor, the characteristics of the established arc are independent of a surrounding atmosphere at normal or low pressures. The short arc is characterized by its constant voltage for currents above a minimum value called the minimum arcing current of the contact. In contrast to the short arc, the long arc between contacts at a fixed separation has a voltage drop which decreases with an increase in current.[12] *Most arcs occurring between contacts on both make and break of telephone switching circuits, are short arcs.* In Table I, Column 3, are given our measured values of the short arc voltage for a few materials.

ARC INITIATION AND TERMINATION CURRENTS

The arc termination current or minimum arcing current is defined as the lowest current at which the arc can be sustained. The arc is extin-

* The arc voltage is about 50 per cent higher except for the carbon arc which has a much higher voltage; see Table I, Column 3.

[12] See for instance: K. Gaulrapp, Untersuchung der elaktrisehen Eigenschaften des Abreisbogens, Ann. Physik. **25**, p. 705, 1936.

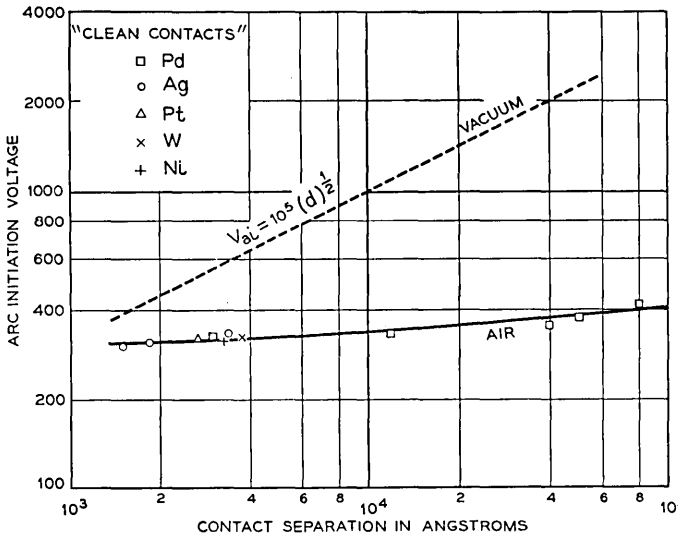


Fig. 4 — Arc initiation voltages for "Clean" metals.

guished when the circuit current drops to this value. To initiate the arc a minimum current must be furnished by the circuit called the arc initiation current. It was previously shown[4] that the arc initiation and termination currents are essentially the same numerically. The existence of these limiting currents as such, rather than current densities, is not understood from a fundamental standpoint.

The limiting currents are a function of the contact material and are appreciably affected by the surface condition of the contacts. Surface contaminations generally reduce the limiting currents of the contacts. The results presented here were obtained by measuring the residual voltage in an R-C circuit following an arc. This voltage is equal to $I_m R$, from which I_m was determined. Our measurements are given in Table I, Column 4 for "clean" contacts. The maximum spread is 20 per cent of the average. For normal contacts, however, surface contamination causes a wide variation in the results. Furthermore, the maximum intensity of the arc, or the maximum current furnished by the measuring circuit, was found to have an appreciable affect on the measurements. This appears to be due to the surface cleaning action of the arc. This effect is demonstrated in Fig. 7 for "normal" palladium contacts. Each point represents an individual measurement of I_m plotted against the corresponding maximum current during the arc. An R-C contact circuit was used. While the measurements show a considerable spread, they indicate a definite trend of an increase in I_m with increasing I_{max}.

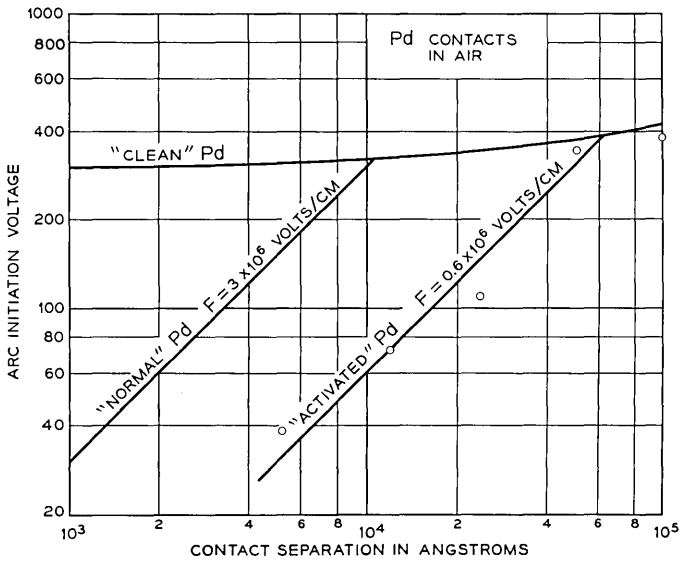


Fig. 5 — Arc initiation voltages for "Activated" palladium contacts.

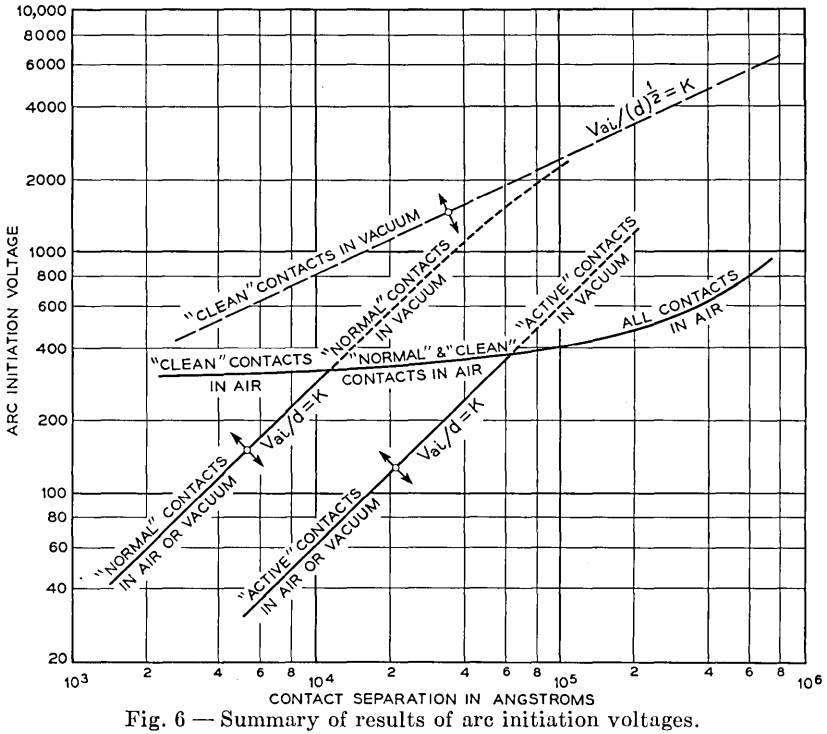


Fig. 6 — Summary of results of arc initiation voltages.

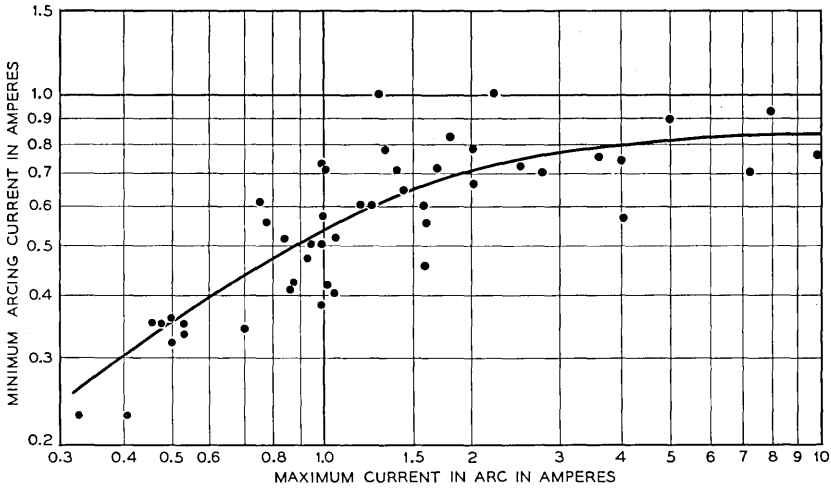


Fig. 7 — Dependence of minimum arcing current on maximum arc current.

In the first part of this paper it was shown that contact activation by organic vapors reduces the arc initiation voltage for a fixed separation. In other words, for a pair of closing contacts the arc will be initiated at a wider separation and a longer arcing time is obtained. In addition, activation tends to decrease the minimum arcing current. Germer[8], measured a minimum arcing current of only 0.027 to 0.037 ampere for active silver. Our measurements for active palladium gave a minimum arcing current of 0.1 ampere. This substantial decrease in the minimum arcing current of contacts due to surface activation usually causes a further increase in the arcing time. Contact activation, therefore, enhances arcing between closing contacts in two ways; first, by initiating the arc at wider separations, and second by maintaining the arc at much smaller currents. The following results are presented to indicate the quantitative significance of contact activation. A pair of "normal" palladium contacts were operated in air saturated with d-limonene at 3 cps. The contacts closed a circuit consisting of a 0.5-microfarad condenser, charged to 50 volts, in series with a 100-ohm resistor. The transient on make was observed on a cathode ray oscilloscope to determine the arcing time. The arc energy E_a was calculated and the ratio $E_a/Cv\ (V_0 - v)$ was plotted against the number of operations, Fig. 8. The denomenator is the maximum arc energy, which is only attained if the arc is maintained until the current reaches zero. The results indicate a rapid increase of the arc energy corresponding to an increase in surface activation. When the contacts become fully active, the arc energy was about two orders of

magnitude greater than the energy for inactive contacts. Contacts may also be activated by inorganic films. Experiments on palladium and silver contacts have shown* that glow discharge between contacts operating in air produced a second type of activation. Nitrides were formed on the contact surfaces and the minimum arcing current dropped to about 0.1 ampere for silver and 0.2 ampere for palladium. This effect was more pronounced with silver contacts.

In carrying out the measurements of the minimum arcing current it was observed that the arc was generally interrupted by one of three causes (1) the minimum arcing current was reached, (2) physical closure of the contacts, and (3) shorting of the arc by a metal bridge formed during the arc. A study of the bridge formation has shown that the height of the bridge was a function of the arc energy. The bridge height was measured by setting the zero separation point before and after the arc. The difference gave the height of the bridge. This was plotted in Fig. 9 against the measured arc energy. The height of the bridge increased roughly with the cubic root of the arc energy up to energies of

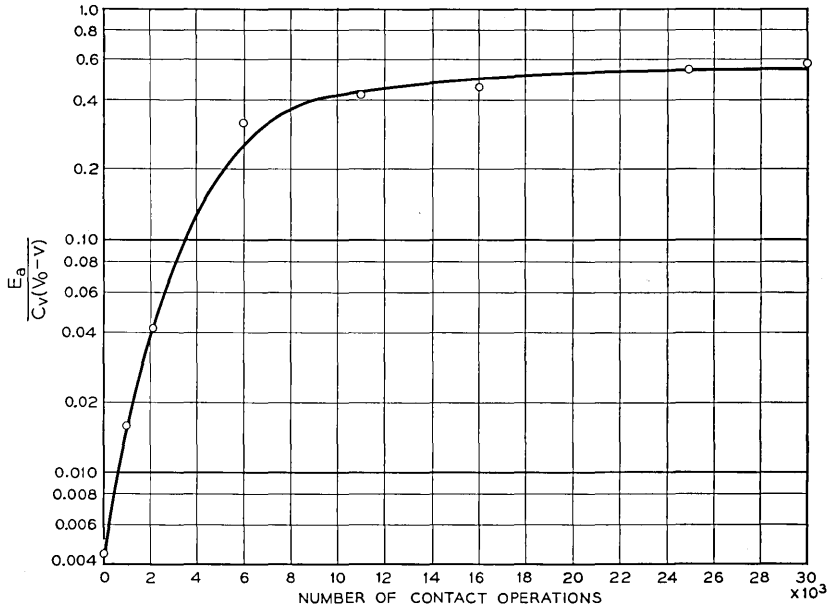


Fig. 8 — Increase in arc energy due to contact "Activation".

* This information was obtained from unpublished work of F. E. Haworth. See also Reference 8 for a discussion of the effects of insulating films on arc initiation.

about 800 ergs. This was followed by a rapid transition into metal *loss* instead of a bridge. This loss increased with increase in energy.

DISSIMILAR CONTACTS — EFFECT OF POLARITY ON ARC CHARACTERISTICS

An experiment was carried out to find the contributions of the anode and cathode in determining the characteristics of the arc. Use was made of the fact that carbon contacts are unique in having low minimum arcing current and high arc voltage compared to most metals; see Table I. Furthermore, carbon contacts always gave a constant field strength line for arc initiation at small separations even when they were cleaned by the heavy arcing process. Tests were carried out with a variety of contact metals against carbon. All the results obtained were qualitatively the same. For illustration only the experiments with palladium-carbon contacts are reported here.

Test specimens were carefully prepared in the following fashion. Palladium to palladium contacts, mounted on the cantilever bar set-up, were cleaned by the heavy arcing process. One contact was removed and replaced by a carbon contact with its surface polished and freed from

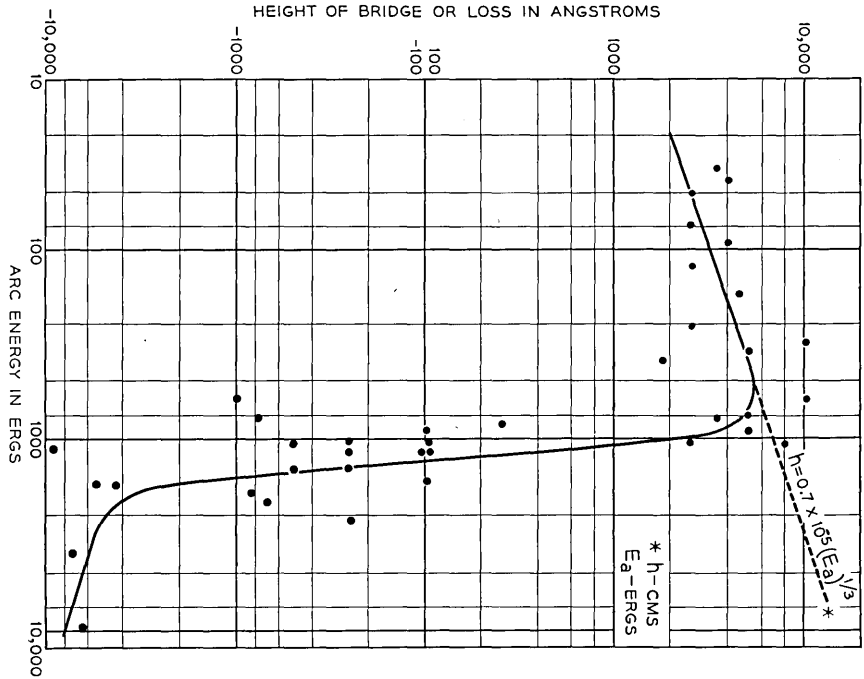


Fig. 9 — Bridge formation during the arc.

TABLE II — EFFECT OF POLARITY ON THE CHARACTERISTICS OF THE ARC — PALLADIUM-CARBON CONTACTS

(1) Contact Configuration	(2) Arc initiation voltage at 6000Å separation	(3) Arc voltage v	Minimum arcing current. I_m amps.
C^+, Pd^-	300	13–15	0.2–0.5
Pd^+, Pd^-	320	14–15	1.1
C^-, Pd^+	130	20–30	0.2–0.3
C^-, C^+	120	20–43	0.03

loose particles. The separation between the contacts was set at 6000Å. By gradually raising the voltage across the contacts and observing it on a cathode ray oscilloscope, the arc initiation voltage was determined. The same measurement was repeated several times. Preceding *each* measurement, the contacts were recleaned by the same process explained above. The polarity was then reversed and a new set of measurements was taken. The results are shown in Table II, Column 2. They indicate that the arc initiation voltage is determined by the cathode. This furnishes support to the postulate that field emission is the first step of the mechanism of arc initiation. By recording the voltage across the contacts during the arc, measurements were made of the arc voltage and the minimum arcing current. The results are given in Table II, Columns 3 and 4. The arc voltage measurements indicate rather conclusively that the arc voltage is determined by the cathode. The minimum arcing current measurements, however, were only slightly, yet consistently, higher with a palladium cathode. It is thought that during a single short arc, particularly with the high intensity arcs used, there is a certain amount of exchange of materials between the electrodes. This exchange is possibly responsible for the observed influence of the anode on the arc characteristics.

It is concluded that the arc initiation voltage as well as the arc voltage are characteristics of the cathode while the minimum arcing current seems to be influenced by both electrodes with stronger inclination towards the cathode characteristics.* The following reservation, how-

* Early experiments by H. E. Ives[13] have led to the conclusion that the arc voltage is a characteristic of the *anode*. In his experiments, however, currents of the order of one ampere were established in the circuit while the *contacts were closed*. Arc measurements were made during the subsequent break of the contacts. It is thought, therefore, that metal bridges must have formed during the break, transferring metal from the anode to the cathode.[2] The arc produced must have been influenced accordingly. In our experiments this difficulty was entirely eliminated.

[13] H. E. Ives, Minimal Length Arc Characteristics, J. Franklin Inst. **198**, No. 4, 1924.

ever, has to be made. All these measurements, although made at a voltage range between 50 and 400 volts corresponding to a contact separation range between 1500 and 100,000Å, had high maximum currents above 1.5 amperes. The energy dissipated in each arc was of the order of thousands of ergs. This reservation is made because recent studies of metal transfer have indicated a reversal in the direction of transfer between the anode and the cathode depending on the rate of energy dissipation in the arc.

ACKNOWLEDGEMENTS

I am indebted to L. H. Germer, P. Kisliuk and F. H. Haworth for much valuable discussion and to A. S. Timms for assistance with many of the experiments reported here.

Abstracts of Bell System Technical Papers* Not Published in this Journal

ANDERSON, J. R.[1]

Electrical Delay Lines for Digital Computer Applications, I.R.E., Trans. P.G.E.C., **2,** pp. 5–13, June, 1953.

A survey of existing lumped parameter and distributed parameter delay lines has shown that their maximum storage capacity is about 23 pulses and 15 pulses respectively regardless of total delay time. An analysis of pulse transmission through distributed delay lines indicates that dissipation in the inductive elements is the chief factor limiting storage capacity. A method is proposed for decreasing this dissipation through the use of high Q nickel zinc ferrites around straight conductors for inductive elements.

ARMSTRONG, C. A.[2]

Communications for Civil Defense, A.I.E.E., Trans., Commun. & Electronics, **7,** pp. 315–326, July, 1953.

BAROTTA, P. J., see S. P. GENTILE.

BIRDSALL, H. A., see D. A. MCLEAN.

BOGERT, B. P.[1]

On the Band Width of Vowel Formants, Letter to the Editor, J. Acoust. Soc. Am., **25,** pp. 791–792, July, 1953.

Measurements were made on a sample of vowel utterances, by male talkers, of the band widths of the first three formants. It was found that the band

* Certain of these papers are available as Bell System Monographs and may be obtained on request to the Publication Department, Bell Telephone Laboratories, Inc., 463 West Street, New York 14, N. Y. For papers available in this form, the monograph number is given in parentheses following the date of publication, and this number should be given in all requests.

[1] Bell Telephone Laboratories.

[2] American Telephone and Telegraph Company.

width was essentially constant and independent of the particular vowel. The mean values for bars 1, 2, and 3 were 130, 100, and 185 cps, respectively. Ten per cent of the 300 band widths measured were less than 90 cps and ten per cent greater than 260 cps.

BRANGACCIO, D. J., see C. C. CUTLER.

BRATTAIN, W. H., see A. M. PORTIS.

BULLINGTON, K.[1]

Frequency Economy in Mobile Radio Bands, I.R.E., Trans., P.G.V.C., **3,** pp. 4–27, June, 1953 (Monograph 2109).

CALBICK, C. J., see D. A. MCLEAN.

CAMPBELL, R. D.[2]

Path Testing for Microwave Radio Routes, Elec. Eng., **72,** pp. 571–577, July, 1953.

CAMPBELL, W. E.[1]

Solid Lubricants, Lubrication Eng., **9,** pp. 195–200, Aug., 1953.

CICCOLELLA, D. F.[1] AND L. J. LABRIE[1]

High Frequency Crystal Units for Use in Selective Networks and Their Proposed Application in Filters Suitable for Mobile Radio Channel Selection, I.R.E., Trans, P.G.V.C. **3,** pp. 118–128, June, 1953.

CONWELL, E. M.[6]

High Field Mobility in Germanium With Impurity Scattering Dominant, Phys. Rev., **90,** pp. 769–772, June 1, 1953.

Experimental measurements show a variation of mobility with electric field intensity of electrons in n type germanium which differs at 20 degrees K from that observed in the same specimen at 77 degrees K and higher temperatures. This difference can be accounted for by scattering by ionized impurities. A crude quantitative treatment is carried out along the lines of Shockley's treatment for the case of lattice scattering. As in that case, the resulting theory fits the data well if the rate of energy loss is taken several times higher than that given by the theory assuming that the surfaces of constant energy are spherical.

[1] Bell Telephone Laboratories, Inc.
[2] American Telephone and Telegraph Company.
[6] Formerly Bell Telephone Laboratories.

CUTLER, C. C.[1] AND D. J. BRANGACCIO[1]

Factors Affecting Traveling Wave Tube Power Capacity, I.R.E., Trans., P.G.E.D., **3,** pp. 9–23, June, 1953.

DACEY, G. C.[1]

Space-Charge Limited Hole Current in Germanium, Phys. Rev., **90,** pp. 759–763, June 1, 1953.

A situation can arise in semi-conductors similar to the space-charge limited emission of electrons in vacuum. The theory of Shockley and Prim for this phenomenon has been extended to the high field case using the approximation that the drift velocity of the carriers is $v = \mu(EE_0)^{1/2}$, where μ is the low field mobility, E the electric field, and E_0 the "critical field." For this approximation the current density analogous to Child's law for a plane parallel diode is

$$J = (\tfrac{2}{3})(\tfrac{5}{3})^{3/2}K\mu E_0^{1/2}V_a^{3/2}/w^{5/2},$$

where V_a is the potential across a diode of thickness w and K is the dielectric constant in mks units. Good agreement between theory and experiment for hole flow in germanium at liquid air temperature has been obtained, using values of μ and E_0 obtained independently by Ryder.

ESHELBY, J. D.[8], READ, W. T.[1] AND W. SHOCKLEY[1]

Anisotropic Elasticity with Applications to Dislocation Theory, Acta Metallurgica, **1,** pp. 251–259, May, 1953.

The general solution of the elastic equations for an arbitrary homogeneous anisotropic solid is found for the case where the elastic state is independent of one (say x_3) of the three Cartesian coordinates x_1, x_2, x_3. Three complex variables $z(\ell) = x_1 + p(\ell)x_2 (\ell = 1, 2, 3)$ are introduced, the $p(\ell)$ being complex parameters determined by the elastic constants. The components of the displacement (u_1, u_2, u_3) can be expressed as linear combinations of three analytic functions, one of $z_{(1)}$, and of $z_{(2)}$, and one of $z_{(3)}$. The particular form of solution which gives a dislocation along the x_3-axis with arbitrary Burgers vector (a_1, a_2, a_3) is found. (The solution for a uniform distribution of body force along the x_3-axis appears as a by-product.) As is well known, for isotropy we have $u_3 = 0$ for an edge dislocation and $u_1 = 0$, $U_2 = 0$ for a screw dislocation. This is not true in the anisotropic case unless the x_1x_2 plane is a plane of symmetry. Two cases are discussed in detail, a screw dislocation running perpendicular to a symmetry plane of an otherwise arbitrary crystal, and an edge dislocation running parallel to a fourfold axis of a cubic crystal.

FULLER, C. S.[1] AND J. A. DITZENBERGER[1]

Diffusion of Lithium Into Germanium and Silicon, Letter to the Editor, Phys. Rev., **91,** p. 193, July 1, 1953.

[1] Bell Telephone Laboratories, Inc.
[8] University of Illinois.

GENTILE, S. P.[1] AND P. J. BAROTTA[4]

Transistor Physics Simplified, Radio and Telev. News, **50,** pp. 44–46, 100–102, July, 1953.

HINES, M. E.[1]

Traveling-Wave Tube, Radio and Telev. News, **49,** pp. 12–14, 26, June, 1953.

KOLB, E. D., see W. P. SLICHTER.

LABRIE, L. J., see D. F. CICCOLELLA.

LINVILL, J. G.[1]

Transistor Negative-Impedance Converters, I.R.E., Proc., **41,** pp. 725–729, June, 1953.

MAY, A. S.[1]

Microwave System Test Equipment, Commun. Eng., **13,** pp. 24–25, 44–45, May–June, 1953.

MCKAY, K. G.[1]

Bombardment Conductivity, Ind. Diamond Rev., **13,** pp. 127–130, June, 1953.

MCLEAN, D. A.,[1] BIRDSALL, H. A.[1] AND C. J. CALBICK[1]

Microstructure of Capacitor Paper, Ind. and Eng. Chem., **45,** pp. 1509–1515, July, 1953 (Monograph 2142).

MCMILLAN, B.[1]

Basic Theorems of Information Theory, Ann. Math. Stat., **24,** pp. 196–219, June, 1953 (Monograph 2124).

This paper describes briefly the current mathematical models upon which communication theory is based, and presents in some detail an exposition and partial critique of C. E. Shannon's treatment of one such model. It then presents a general limit theorem in the theory of discrete stochastic processes, suggested by a result of Shannon's.

[1] Bell Telephone Laboratories, Inc.
[4] Hudson Technical Institute.

MERTZ, P.[1]

Influence of Echoes on Television Transmission, J.S.M.P.T.E., **60,** pp. 572–596, May, 1953 (Monograph 2144).

PETERSON, G. E.[1]

Basic Physical Systems for Communication Between Two Individuals, J. Speech and Hearing Disorders, **18,** pp. 116–120, June, 1953 (Monograph 2135).

PIERCE, J. R.[1]

Transistors, Radio-Electronics, **24,** pp. 42–44, June, 1953.

PORTIS, A. M.,[7] A. F. KIP,[7] C. KITTEL,[7] AND W. H. BRATTAIN[1]

Electron Spin Resonance in a Silicon Semi-Conductor, Letter to the Editor, Phys. Rev., **90,** pp. 988–989, June 1, 1953.

PRIM, R. C., see W. SHOCKLEY.

QUARLES, D. A.[5]

Progress and Problems, Elec. Eng., **72,** pp. 667–669, August, 1953.

QUARLES, D. A.[5]

Report to the Membership, Elec. Eng., **72,** pp. 477–479, June, 1953.

READ, W. T., see J. D. ESHELBY.

RYDER, E. J.[1]

Mobility of Holes and Electrons in High Electric Fields, Phys. Rev., 90, pp. 766–769, June 1, 1953.

The field dependence of mobility has been determined for electrons and holes in both germanium and silicon. The observed critical field at 298 degrees K beyond which μ varies as $E^{-1/2}$ is 900 volts/cm for n-type germanium, 1400 volts/cm for p-type germanium, 2500 volts/cm for n-type silicon, and 7500 volts/cm for p-type silicon. These values of critical field are between two to four times those calculated on the basis of spherical constant energy surfaces in the Brillouin zone. A saturation drift velocity of $6(10)^6$ cm/sec is observed in germanium which is in good agreement with predictions based on scattering

[1] Bell Telephone Laboratories, Inc.
[5] Sandia Corporation.
[7] University of California, Berkeley.

by the optical modes. Data on n-type germanium at 20 degrees K show a range over which impurity scattering decreases and the mobility increases with field until lattice scattering dominates as at the higher temperatures.

SHOCKLEY, W., see J. D. ESHELBY.

SHOCKLEY, W.[1] AND R. C. PRIM[1]

Space-Charge Limited Emission in Semi-Conductors, Phys. Rev., **90,** pp. 753–758, June 1, 1953.

A situation analogous to thermionic emission into vacuum can occur in semi-conductors. A semi-conductor analog for a plane parallel vacuum diode may consist of two layers of n type semi-conductor bounding a plane parallel slab of pure semi-conductor. The current density analogous to Child's law is $J = 9_{\kappa\epsilon_0\mu}V^2/8W^3$, where κ = dielectric constant, ϵ_0 = mks permittivity, μ = mobility, V = applied voltage, and W = thickness of pure region. The condition prevailing at the space-charge maximum is analyzed taking into account diffusion due to random thermal motion. Brief discussions are given of the effect of fixed space charge, the dependence of mobility upon electric field strength and the role of space-charge limited emission in a new class of unipolar transistors.

SLICHTER, W. P.[1] AND E. D. KOLB[1]

Solute Distribution in Germanium Crystals, Letter to the Editor, Phys. Rev., **90,** pp. 987–988, June 1, 1953.

TOWNSEND, J. R.[1]

A Dynamic Program for Conversion, Metal Progress, **63,** pp. 79–81, June, 1953.

TURNER, E. H.[1]

New Non-Reciprocal Waveguide Medium Using Ferrites, Letter to the Editor, I.R.E., Proc., **41,** p. 937, July, 1953.

WANNIER, G. H.[1]

Threshold Law for Single Ionization of Atoms or Ions by Electrons, Phys. Rev., **90,** pp. 817–825, June 1, 1953.

When an electron hits an atom or ion, it may knock off an electron. This process is fundamental in almost all types of gas discharge. The reaction is endothermic; hence there is a threshold value in the electron energy below

[1] Bell Telephone Laboratories, Inc.

which it does not occur. In this paper, the dependence of the yield on the energy just above this threshold is derived. The derivation is not rigorous because it circumvents some of the difficulties of the three-body problem by applying ergodicity, albeit in a weakened form. The result is that, for atoms, the yield rises as the 1.127th power of the energy excess. For ions the exponent lies between this number and unity.

WILLARD, G. W.[1]

Ultrasonically Induced Cavitation in Water: A Step-by-Step Process, J. Acoust. Soc. Am., **25,** pp. 669–686, July, 1953.

A 2.5-mc, barium-titanate, spherically focusing radiator was used to produce cavitation in both degassed and aerated water entirely within the restricted, high intensity focal region, remote from the water boundaries. The sonic intensity rises to 1.8 kw/cm^2 and the pressure amplitude to ± 70 atmospheres at the focus. High-intensity illumination and an unusual high speed photographic technique permit observation and timing of the step-by-step process of cavitation development.

Feather-shaped cavitation bursts are sporadically produced, being initiated in the insignificant quill portion nearest the radiator, then abruptly expanding to form the catastrophic plume portion. The plume is believed to be formed by myriads of microcavities, too small and close for individual observation. These two fundamental steps are identically produced, and with equal ease, both in degassed and aerated water. The whole action is over in several milliseconds, except that in the case of aerated water a third bubble step is produced. In aerated water, non-collapsing gas bubbles are generated by and concurrently with, the catastrophic step. These bubbles remain after collapse of the burst, to be blown off down stream by the sonically induced liquid streaming.

The bubble step is not generated without the presence of the catastrophic step. The latter is generated only if the initiation step reaches a definite degree of development (not always attained). This requires sonic activation for increasing lengths of time for decreasingly smaller sonic intensities. Origination of the initiation step, and hence of the whole cavitation phenomena, is believed to occur whenever a stray nucleus (weak spot) streams into the high intensity sonic field.

YOUNG, W. R.[1]

Comparison of Mobile Radio Transmission at 150, 450, 900 and 3700 Mc, I.R.E., Trans., P.G.V.C. **3,** pp. 71–83, June, 1953.

[1] Bell Telephone Laboratories, Inc.

Contributors to this Issue

M. M. ATALLA, B.S., Cairo University, 1945; M.S., Purdue University, 1947; Ph.D., Purdue University, 1949; Studies at Purdue undertaken as the result of a scholarship from Cairo University for four years of graduate work. Bell Telephone Laboratories, 1950–. For the past three years he has been a member of the Switching Apparatus Development Department, in which he is supervising a group doing fundamental research work on contact physics and engineering. Current projects include fundamental studies of gas discharge phenomena between contacts, their mechanisms, and their physical effects on contact behavior; also fundamental studies of contact opens and resistance. In 1950, an article by him was awarded first prize in the junior member category of the A.S.M.E. He is a member of Sigma Xi, Sigma Pi Sigma, and Pi Tau Sigma, and a junior member of the A.S.M.E.

FRANK E. BLOUNT, B.S. in E.E., Oregon State College, 1928. Bell Telephone Laboratories 1928–. Mr. Blount tested panel system circuits for a year and then transferred to development work on special switching and signaling circuits. His interest in circuit design has also included circuits for automatic toll ticketing for the step-by-step system, for radar testing, and more recently for No. 5 crossbar system. Member of Tau Beta Pi, Eta Kappa Nu and Phi Kappa Phi.

J. T. LINDSAY BROWN, B.S. College of the City of New York, 1915. Western Electric Company, 1915–1925. Bell Telephone Laboratories 1925–. For twenty-five years Mr. Brown was concerned with the development and testing of telephone instruments and such allied apparatus as loudspeakers, microphones, and earphones. In 1940 he transferred to work on the development of glass sealed magnetic switches. He is currently in charge of a group developing mercury contact relays. Member of the A.I.E.E., I.R.E. and Acoustical Society of America.

WALLACE A. DEPP, B.S. and M.S. in E.E., University of Illinois, 1936 and 1937. Bell Telephone Laboratories, 1937–. In his early laboratories

association Mr. Depp was concerned with thoriated tungsten and tantalum emitters and later with cold cathode tubes. During World War II he worked on pulsing thyratrons and fixed spark gap tubes for radar, and miniature thyratrons used in the proximity fuse. He was subsequently in charge of the basic development of all types of gas-filled tubes. Recently he transferred to Transmission Systems Development with responsibility for broad band carrier terminal equipment, N and O carrier systems and automatic switching for the L3 coaxial cable system. Member of the A.I.E.E., Eta Kappa Nu, Tau Beta Pi, Phi Kappa Phi and Sigma Xi. Senior member of the I.R.E.

JAMES M. EARLY, B.S. cum laude, New York State College of Forestry, 1943; M.S. and Ph.D. Ohio State University, 1948 and 1951. Bell Telephone Laboratories 1951–. After teaching Electrical Engineering at Ohio State University for five years while studying for his Master's and Ph.D. degrees, Dr. Early joined an electronic apparatus development group, participating in the development of the junction transistor. At present he is doing theoretical as well as development work on high frequency junction transistors. Member of the I.R.E. and Eta Kappa Nu. Associate of Sigma Xi.

JOSEPH GRAMELS, B.S. in E.E., New York University, 1936. Bell Telephone Laboratories 1925–. Mr. Gramels was first occupied with testing and development work in transmission, including handsets, recording apparatus and 33 rpm records. In 1937 and 1938 he worked on electrolytic condensers and silicon carbide varistors. Since 1938 he has been concerned with investigations of selenium rectifier cells both for Bell System applications and for military use. Member of the A.I.E.E.

MASON A. LOGAN, B.S. in Physics and Engineering, California Institute of Technology, 1927; M.A. in Physics, Columbia University, 1933. Carnegie Institute of California, Seismological Laboratory, 1926–1927. Bell Telephone Laboratories, 1927–. His early Laboratories' projects were concerned with wire transmission problems particularly those of losses, noise and cross induction in local, manual and dial telephone circuits. This was followed by circuit research on alternating current methods of signaling including the use of non-linear elements and electronic terminal equipment. From 1941 to 1948 he worked on military projects, including a mine fire control system, anti-aircraft gun director, magnetic proximity fuses, and guided missiles. For the past five years he has been a

member of the Switching Apparatus Development Department in which he is supervising a group concerned with static and dynamic behavior of new electromagnets and relays. He is also engaged in investigations of the performance of electrical contacts on telephone relays.

C. E. POLLARD, JR., Polytechnic Institute of Brooklyn, 1927–1931. Bell Telephone Laboratories, 1925–. Mr. Pollard first worked on voice recording and reproducing equipment and spent some time on the development of telephone carbon microphones before becoming interested in mercury contact relays. In this field he has been concerned with a wide variety of relays, and during World War II concentrated on their application to military projects. He is currently engaged in development work on mercury contact relays.

JOHN H. ROWEN, B.E.E., Ohio State University, 1948; M.S., Ohio State University, 1951. U.S.N.R., maintenance of Air Force radar equipment, 1944–1946; Ohio State University, Research Foundation, Antenna Laboratory, 1948–1951; Bell Telephone Laboratories, 1951–. Concerned with the practical application of fundamental research, he has spent his two years at Bell Laboratories on studies of the microwave behavior of ferrites. While at Ohio State University, Mr. Rowen worked on several methods of measuring the radiation efficiency of small aperture antennas. Member of the I.R.E. and Eta Kappa Nu.

MARK A. TOWNSEND, B.S. in E.E., Texas Technological College, 1936; S.M. in E.E., Massachusetts Institute of Technology, 1937. General Electric Company, 1937–1943; Massachusetts Institute of Technology, Radar School, 1943–1945. Bell Telephone Laboratories, 1945–. Mr. Townsend has been concerned with the basic development of gas-filled tubes. His projects have included work on the voltage reference tube, cold cathode stepping tubes, and the development of tubes for use in transmission and switching. At present Mr. Townsend is in charge of a group responsible for basic development of gas-filled tubes. Member of the A.I.E.E., Tau Beta Pi and Sigma Xi.