George A. Campbell, 1870-1954

The passing of Dr. George Ashley Campbell November 10, 1954, marks the close of an era that has seen the communication art grow from infancy to robust manhood.

The problems of telephone transmission were little understood when

1

George Campbell, a young mathematical physicist joined the American Bell Telephone Company (later American Telephone and Telegraph Company) in 1897. He had just returned from four years of study at various universities in Paris, Vienna, and Göttingen, having received his B.S. degree in 1891 from Massachusetts Institute of Technology, A.B. from Harvard, 1892, and M.A. in 1893 from the same institution. In these early years at American Bell, he also continued his studies at Harvard, where in 1901, he received the Ph.D. degree.

Familiar with the work of Rayleigh and Heaviside, Campbell's first undertaking was to find some method of mitigating the attenuation of voice currents in telephone lines, which theretofore, produced a barrier against telephone communication over very long distances. Heaviside had shown that inductance, if properly applied in a long telephone circuit, should diminish rather than increase the attenutation. Campbell followed this suggestion and developed a theory of "loading," but there occurred one of those rare coincidences in the history of science of two independent investigators arriving at substantially the same result at the same time. Without knowing that, at Columbia University, Professor M. I. Pupin was studying the same problem, Campbell applied for a patent at about the same time as Pupin. After final adjudication under the technical rules of the patent law, the basic patent was awarded to Pupin, on the ground of priority of conception. As Dr. F. B. Jewett so aptly said, "It should be mentioned, however, that Campbell's analysis of the problem — actually more detailed than Pupin's — led him to formulate rules for the design of loading coils and their spacing, which were, from the very beginning, the only ones employed in this country. By this one piece of work performed within a relatively short time after his employment in the Bell Company, Campbell demonstrated his unique ability at mathematical physics, as well as his knack for stating conclusions in a form that the development engineer could use in practical applications."

As a consequence of his work on loaded lines — indeed, almost as a part of it — Dr. Campbell arrived at the idea of an electric wave filter. By a suitable network of coils and capacitors, he was able to produce a device that would allow a preferred band of frequencies to pass through it — such as those of the human voice, for example — while discriminating markedly against frequencies above and below the desired band. This development was first used in so-called "carrier" systems over long distance open-wire lines, which increased by several fold the number of telephone circuits carried by them. The principle became of even greater importance when much wider bands of frequencies were employed in

the coaxial cable and in the newer forms of radio transmission designed to transmit, over the same electrical path, either many hundreds of telephone conversations, or several television programs. It will also be used in the new transatlantic telephone cable.

As the effectiveness of telephone instruments increased and the length of circuits grew, "crosstalk," the tendency for conversation in one circuit to be heard in another, became an obstacle to telephone advance. Earlier workers had shown that this crosstalk was a complex effect resulting partly from electromagnetic induction and partly from electrostatic induction. Campbell turned his attention to this problem between 1903 and 1907, and showed that crosstalk between two circuits depends to a considerable extent, and particularly for loaded circuits, on the direct capacitance between the wires of the two circuits. He termed this function the "Direct Capacity Unbalance." This work led not only to mathematical formulae, but to the development of measuring apparatus which was destined to play a great part in future telephone developments. In this period he produced his well-known shielded balance for the accurate measurement of electrical constants at telephone frequencies. This contributed greatly to the development of cables to replace open-wire lines. Out of it also grew the whole shielding technique applied today in innumerable ways in the high-frequency art.

Closely associated with his work on filters and line balance were problems concerning telephone repeater circuits. Early in telephone development, various experimenters and inventors had proposed circuits to accomplish two-way telephony, and from this prior work had come two fundamental repeater circuits, one in which a single repeating element amplifies messages reaching it from both directions, and one that includes two amplifier elements, one assigned to each direction of transmission. In the first, two sections of line, as nearly identical as possible, are balanced against each other as opposite arms of a bridge. In the second, each incoming section of line is balanced against an artificial line or network, thus permitting, as Campbell's analysis showed, a greater inherent flexibility as well as a greater stability. Campbell's studies closely formulated the stability limits of the circuits, and in addition, led to the use of the "four-wire circuit" as a logical extension of the one-way paths in the second type repeater, each extended path containing as many one-way amplifiers and line sections as desired. Although when it was proposed in 1912 it was looked upon as having little practical application, its later wide use at both voice frequencies and in the carrier art, shows it to be an example of a technical advance that could not be employed until a complex art had caught up with it.

As a mathematician, he was interested not only in solving problems, but also in the logical relations of mathematics itself; perhaps his most significant contribution in this field was in the table of Fourier Integrals for practical applications in the study of transients and other non-periodic phenomena.

In this brief review of Dr. Campbell's work, mention can be made of only a few of his many contributions. Those that have been chosen are representative of his more outstanding contributions to electrical communication. His career was highly productive of discoveries, inventions, and patents. His achievements entitle him beyond question to rank first among his generation of theoretical workers in electrical communication. In recognition of his distinguished contributions, he was awarded the Medal of Honor of the Institute of Radio Engineers in 1936, the Elliott Cresson Gold Medal of the Franklin Institute in 1939, and the Edison Medal of the American Institute of Electrical Engineers in 1940.

At the time of his retirement, he was a member of the American Mathematical Society, American Mathematical Association, American Physical Society, The American Association of Physics Teachers, American Academy of Arts and Sciences, and the Harvard Engineering Society. He was also a Fellow of the American Institute of Electrical Engineers and of the American Association for the Advancement of Science. In appreciation of Dr. Campbell's long and distinguished service and of his fundamental contributions to the development of electrical communication, his technical papers were collected and published by the American Telephone and Telegraph Company, in a bound volume entitled "Collected Papers of George A. Campbell" (1937).

Though Campbell's achievements are now best known by his personal mathematical contributions, he contributed to the advance of telephony in another way that should not be forgotten. He it was who hired Frank B. Jewett and Edwin Colpitts, and others, whose ability and training fitted them for leadership in the application of scientific knowledge and methods to the practical problems of telephony.

He was a gentle and retiring man, but to those who knew him best he was a lovable companion who sought, not fame or high position, but understanding. His interest in the technical problems of telephony continued to the end of a long and active life. Out of his work came useful and readily available mathematical tools for those who followed him, but more importantly, a spirit of inquiry that left its enduring imprint on the character of the Bell Telephone System as we know it today.

Behavior and Applications of Ferrites in the Microwave Region

By A. G. FOX, S. E. MILLER, and M. T. WEISS

The behavior and applications of ferrites in microwave circuits have been studied at the Holmdel Radio Laboratory with particular emphasis on non-reciprocal properties. There are numerous ways of building non-reciprocal devices which accomplish the same function, and the authors propose at the outset that a terminology based on function be adopted. Definitions of the gyrator, circulator and isolator are given.

A qualitative method of deducing the properties of new ferrite-loaded circuits, termed "point-field" analysis, has been particularly fruitful, in that it permitted the exploration of a wide variety of ferrite-loaded circuits despite the absence of precise mathematical analyses.

Faraday rotation in longitudinally-magnetized ferrite-loaded circuits has been studied; the effects of higher-order modes, the reasons for variation in Faraday rotation as a function of frequency, and the optimum geometry for the ferrite loading are discussed. In transversely-magnetized ferrite-loaded rectangular waveguide it is shown that the single mode medium may be non-reciprocal as to phase constant, attenuation constant, or distribution of field components in the cross-section. The latter effect is called the "field-displacement" effect. In transversely-magnetized ferrite-loaded round waveguide, either reciprocal or non-reciprocal birefringence may be achieved and examples are given. Some non-reciprocal ferrite-loaded dielectric waveguides are discussed. In all of these forms of non-reciprocal transmission media, gyrators, circulators and isolators may be built, and experimental data are reported on some of the possibilities. A brief review of prospective uses for these non-reciprocal elements is given.

CONTENTS

1. INTRODUCTION

In the history of the communication art there has never been available any passive circuit element or waveguide medium which was not recip-rocal. After Polder,[1] Beljers,[2] and Roberts[3] demonstrated the gyromag-netic nature of ferrites at microwave frequencies, passive non-reciprocal devices for the first time became possible. Such non-reciprocal devices have great immediate and potential value in the microwave art. Addi-tionally, ferrites have reciprocal properties which are under the control

of an applied magnetic field and which lead to new forms of modulators, variable attenuators, and variable reciprocal phase-changers in the microwave range.

Since C. L. Hogan*,[4] demonstrated the practical importance of ferrites to the microwave art by making use of microwave Faraday rotation to build a gyrator, the authors have studied, at the Holmdel Radio Laboratory, the behavior and applications of ferrites with emphasis on their non-reciprocal properties. It soon became apparent that ferrites can manifest non-reciprocal properties in a variety of ways other than by Faraday rotation. Thus, they may produce non-reciprocal birefringence, non-reciprocal phase shift, non-reciprocal field displacement, non-reciprocal coupling through apertures, as well as non-reciprocal rotation of polarization. Any one of these non-reciprocal properties can be used, together with conventional reciprocal circuit elements, to obtain all of the non-reciprocal circuit functions which have so far been devised. It therefore seems desirable, before proceeding further, to identify some of the more important non-reciprocal circuit elements with individual names and circuit symbols irrespective of the detailed means used to obtain these circuit functions.

1-1. *Definitions and Symbols for Non-Reciprocal Devices*

The *gyrator* is defined[5] as a transmission element having no phase shift for wave propagation through it in one direction, but having a phase shift of π radians for propagation in the opposite direction. A single line circuit symbol for this element is shown in Fig. 1. Lines 1 and 2 represent input and output waveguides. The symbol $\overleftrightarrow{\pi}$ in the box indicates a phase delay of π radians for waves passing in the direction of the arrow, and no phase delay in the reverse direction of propagation. It is of course possible to add any amount of reciprocal phase shift in series with this element so that for one direction of propagation the phase delay may be θ, and for the other direction of propagation the phase delay may be $\pi + \theta$. However, this additional reciprocal phase delay is non-essential, and the function defined by Fig. 1 is the essence of the gyrator behavior. Various practical embodiments are discussed later in this paper.

It may be noted that Tellegen[5] chose to use the *gyrator* as an essential building block, out of which various non-reciprocal circuits could be constructed. For this reason, there has been a tendency for many people to call any non-reciprocal ferrite device a gyrator. We strongly feel that

---

* Contemporary with Hogan's disclosure, C. H. Luhrs marketed a microwave switch employing Faraday rotation in a ferrite-loaded cavity.

this loose usage should be avoided, since many of the structures which have been called gyrators are in no sense the kind of thing which Tellegen defined. Some of these devices may be of more practical importance than the gyrator and are just as capable of being employed as fundamental building blocks. We shall therefore restrict the usage of the term *gyrator* to that defined above in agreement with Tellegen's original definition.

A second circuit element of considerable fundamental importance is the *isolator*, so named because it can be used to isolate one transmission element from reflections arising from succeeding elements. Fig. 2 is a circuit symbol which demonstrates this behavior. The symbol ←⎍ in the box indicates absorption of power in an internal load for propagation from 2 to 1, and a transmission with no absorption of power from 1 to 2. This representation reminds one that a device which prevents transmis-
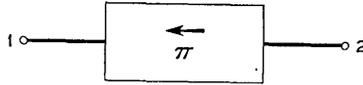
Fig. 1 — Gyrator symbol.
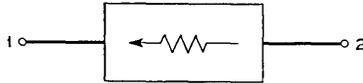
Fig. 2 — Isolator symbol.

sion in only one direction must always do so by absorption rather than by reflection of power, as a consequence of the second law of thermodynamics.

Figure 3(a) shows the circuit symbol for a *circulator*. The term *circulator* was chosen to connote a commutation of power from one transmission-mode terminal to another. Thus, power delivered at terminal 1 passes through the circulator in the direction indicated by the curved arrow to the next terminal, 2, where it emerges. Power inserted at 2 proceeds on around to terminal 3; power inserted at terminal 3 proceeds to terminal 4; and power inserted at terminal 4 finally returns to terminal 1. Although the four-arm circulator is displayed in Fig. 3(a), it is also possible to build three-arm circulators as well as various other multi-arm circulators, as illustrated in Fig. 3(b). Practical means for obtaining these devices are described later in this paper.

Finally, one should note that by employing any of the various means available for providing non-reciprocal phase delay, it is possible to obtain

an arbitrary number of degrees of non-reciprocal phase shift as shown symbolically in Fig. 4. Such a device can be called a *directional phase shifter*. Clearly, the gyrator is merely a special case of the more general *directional phase shifter*.

By clearly distinguishing between these various important circuit functions, both symbolically and in terminology, much confusion can be avoided.

The above mentioned non-reciprocal functions were first realized by means of Faraday rotation in ferrite-loaded round waveguides.* It became apparent at an early stage that the uniform plane-wave explanation of Faraday rotation is totally inadequate to explain the behavior of ferrites in waveguides except in a qualitative way. Measurements of the reflection and transmission properties of ferrite-loaded waveguides show many anomalies which would not be predicted from the infinite uniform

(a)

(b)
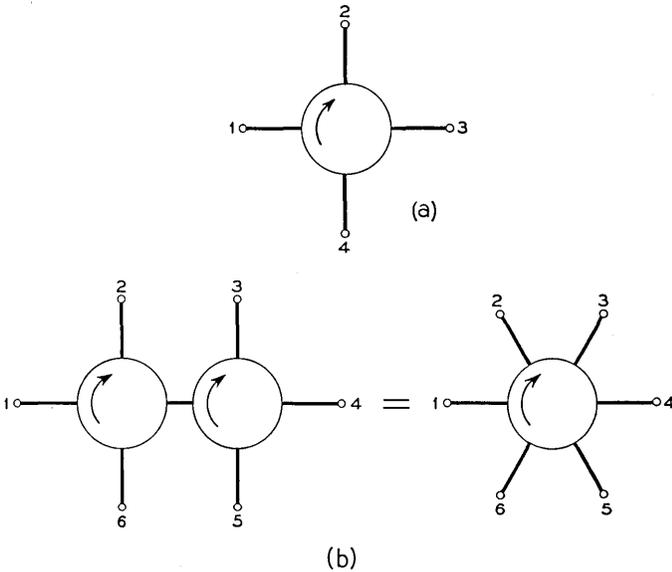
Fig. 3(a) — Circulator symbol.
Fig. 3(b) — Multi-arm circulator symbol.
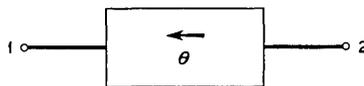
Fig. 4 — Directional phase shifter symbol.

* See Reference 4, C. L. Hogan.

plane-wave theory. On the other hand, a rigorous solution of Maxwell's equations presents formidable mathematical problems which have so far been solved for only a few cases. It is possible, however, to obtain a qualitative explanation of the observed facts together with a considerable amount of physical insight by making use of a "point-field analysis" which is based on some reasonably intelligent guesswork as to the nature of the electromagnetic-wave field patterns. We assume that the ferrite-loaded waveguide contains a perturbed wave bearing a strong resemblance to one or more of the well-known modes in an empty waveguide. We determine the fields which this wave would produce in every part of the ferrite. The behavior of the ferrite at every point can then be predicted on the basis of Polder's tensor permeability theory, or on the basis of the clockwise and counterclockwise permeability theory. Finally, the total behavior can be deduced as a weighted sum of the behavior of all the parts of the ferrite.[6] Not only does this approach give explanations for the observed behavior of Faraday rotation elements, but it has aided greatly in understanding and predicting the behavior of ferrites in radically different geometries making use of transverse applied magnetic fields. It should be emphasized that the nonreciprocal properties of ferrites stem from the gyroscopic nature of the electrons which may be manifested in a variety of ways of which Faraday rotation is but one. In this paper we attempt to build up a physical picture of the way ferrites actually interact with waveguide waves and to show how these same physical concepts have been useful in deriving new non-reciprocal devices.

2. MICROWAVE PROPERTIES OF FERRITES

The magnetic properties of ferrites are essentially the same as those of iron and other ferromagnetic metals. The one outstanding characteristic which distinguishes good microwave ferrites from the magnetic metals and which gives rise to their unique microwave behavior is their extremely high resistivity (typically from 10^{+6} to 10^{+8} ohm-cm for ferrites relative to 10^{-5} for iron). Whereas in the case of iron, a radio wave sees an effective reflector, in the case of ferrite the wave can enter and pass through substantial amounts of the material without excessive reflection or attenuation. In the process, the wave has an opportunity for strong interaction with the spinning electrons which are responsible for the magnetic properties of the material. As a result of this interaction Faraday rotation and other non-reciprocal properties may be manifested.

While the ferrites exhibit a favorably high resistivity, they also have

moderately high dielectric constants at microwave frequencies, which are typically in the range from 10 to 20. This gives rise to important problems in applying ferrites. The high dielectric constant can cause appreciable reflection of power at an air-ferrite boundary, and such reflections may be but slightly reduced by the polystyrene matching cones which have sometimes been used. Also, the presence of high dielectric constant material within a normally single-mode waveguide may give rise to mode conversion (conversion of power from the desired to undesired waves in the ferrite-loaded section) even in the absence of an applied magnetic field. Both of these effects can greatly alter the reciprocal and non-reciprocal transmission properties of the ferrite-loaded region, and many of the capricious results obtained in early studies of ferrites are attributable to reflection and mode conversion difficulties.

The vital property of ferrites which makes them so useful in the microwave art is their unique permeability. There are two ways of defining ferrite permeability — one, in terms of a tensor permeability for linearly polarized waves, and the other, in terms of a scalar permeability for circularly polarized waves.

The ferrite's microwave permeability is due to the effects of certain electrons, which behave en mass gyroscopically according to the classical picture of Fig. 5. It is found that the charge, mass and spin of these electrons are associated with an angular momentum and a magnetic moment in the directions shown, which it may be noted are the same as those to be expected for a spinning positive mass and negative charge.
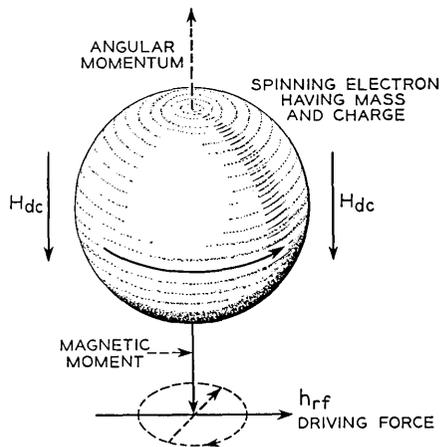
Fig. 5 — Schematic representation of an electron.

With the application of a dc magnetic field (H_{dc}), the axis of the electron spin is aligned with the dc field. If the spin axis is momentarily deflected from parallelism with the dc field, it will not return to its original position immediately but will precess as a gyroscope about H_{dc} at a frequency which is proportional to the magnitude of H_{dc}. This frequency we call the gyromagnetic resonance frequency, and the behavior of ferrites is profoundly dependent upon whether the frequency of applied radio waves is greater or less than the gyromagnetic resonance frequency. If an alternating magnetic field (h_{rf}) is applied perpendicular to H_{dc}, the resulting precession will cause the tip of the magnetic moment vector to describe an elliptical path as indicated by the dashed line in Fig. 5. The direction of rotation around the ellipse will always be clockwise when looking along H_{dc}. As a result, the electron will produce components of magnetic flux perpendicular to both H_{dc} and h_{rf}.*

2-1. *Tensor and scalar permeabilities*

With this classical model as a postulate, Polder[1] has shown that a uniformly magnetized and saturated region of ferrite subjected to uniform rf fields will contain uniform flux densities given by

$$b_x = \mu h_x - jk h_y$$
$$b_y = \mu h_y + jk h_x \tag{1}$$
$$b_z = \mu_0 h_z$$

where

$$\mu = \mu_0 - \frac{\gamma M \omega_0}{\omega_0^2 - \omega^2} \tag{2}$$

$$k = \frac{\gamma M \omega}{\omega_0^2 - \omega^2} \tag{3}$$

---

* This description is a very much simplified explanation of ferrite behavior. For example, most of the electrons in the ferrite are paired with their spins pointing in opposite directions, and these do not contribute to any first order magnetic effects. By virtue of certain forces between them known as "exchange forces," some electrons prefer to line up with their spins in parallel and may be easily oriented en mass by the application of a relatively small magnetic field. This is what makes the material appear "magnetic." At the same time the presence of thermal energy has an opposite effect in that it tends to create disorder in the alignment of electron spins. As a result, the spins never all line up perfectly with the applied field. Nevertheless, we can assume that out of the total number of unpaired electron spins a certain fraction do line up with the applied field and behave in accordance with the simple picture given above for a single electron.

$$\omega_0 = -\gamma H_i \tag{4}$$

μ_0 = permeability of free space

γ = ratio of magnetic moment to angular momentum for the electron

H_i = internal dc field directed along $+ Z$ axis

The rationalized MKS system is used here and throughout this paper. It should be understood that H_i is the macroscopic dc H in the medium calculable by standard magnetostatic techniques. It is not the H which would occur in a spherical cavity in the medium. The components h_x, h_y, and h_z are the total internal rf field components in the macroscopic sense. They include not only the externally applied rf field, but also any demagnetizing fields which may be present. M is the magnetic polarization density and is taken as $B - \mu_0 H$. ω_0 may be interpreted as the natural precession frequency of an isolated electron placed in the steady field H_i. Finally, it is important to note that the magnetic moment points in the opposite direction from its angular momentum. Therefore γ is negative, and a minus sign must be used when substituting in the above equations. Consistent with equation (4), γ must have the dimensions of ω/H. It is equal to

$$0.035 \ \frac{\text{mc}}{\text{amp/meter}}, \text{(or 2.8 mc/oersted)}$$

The tensor form of the equations (1) indicate that an rf h vector applied along the X axis will give rise to rf components of b along both the X and Y axes. Consequently, the k component of the permeability tensor plays the part of a coupling coefficient which can account for the transfer of rf wave power from one polarization to an orthogonal polarization.

An alternative expression for permeability can be derived from the tensor form. Let us assume that the ferrite is excited by a clockwise circularly polarized field vector as seen looking along the dc magnetic field, and this will be designated as a $(+)$ field vector. Then

$$h_y = -jh_x\} \quad (+) \text{ rf } h \text{ vector} \tag{5}$$

Substitution of these in equation (1) gives:

$$\left.\begin{array}{l} b_x = (\mu - k)h_x \\ b_y = -j(\mu - k)h_x \end{array}\right\} \quad (+) \text{ rf } b \text{ vector} \tag{6}$$

Thus, the resultant b is circularly polarized in a clockwise sense. If we now assume a counterclockwise field vector looking along the dc magnetic

field and denoted by $(-)$,

$$h_y = +jh_x\} \; (-) \text{ rf } h \text{ vector} \tag{7}$$

Substitution in equation (1) gives:

$$\left.\begin{array}{l} b_x = (\mu + k)h_x \\ b_y = +j(\mu + k)h_x \end{array}\right\} \; (-) \text{ rf } b \text{ vector} \tag{8}$$

The resultant b is now polarized in a counterclockwise sense. We can see that for circularly polarized fields the total permeability is a scalar. Thus,

$$\mu_+ = \frac{b_+}{h_+} = (\mu - k) = \mu_0 - \frac{\gamma M}{\omega_0 - \omega} \tag{9}$$

$$\mu_- = \frac{b_-}{h_-} = (\mu + k) = \mu_0 - \frac{\gamma M}{\omega_0 + \omega} \tag{10}$$

Fig. 6 illustrates schematically how the μ_+ and μ_- vary as a function of H_i. The μ_+ has been plotted to the right and the μ_- to the left of the $H_i = 0$ axis, with H_i increasing in both directions. While Polder's equations were developed without any losses included, the permeability curve
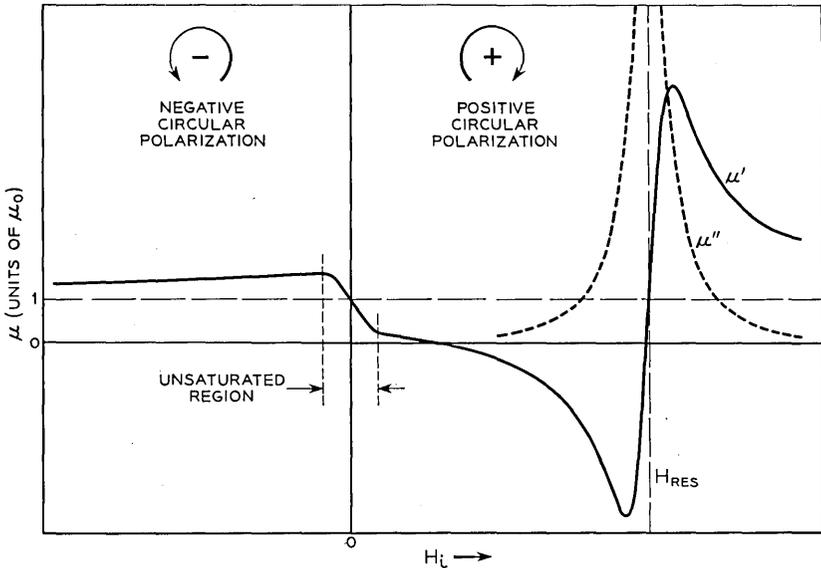
Fig. 6 — Ferrite permeability for circularly polarized waves as a function of internal magnetic field.

is plotted in Fig. 6 as it would be expected to appear if some magnetic loss were present. Thus, the real part (μ') of the permeability remains finite throughout the resonance region. The imaginary or loss component of the permeability (μ'') should rise to a maximum at resonance as shown by the dashed curve. Polder assumed that the ferrite was completely saturated, in which case the magnetization M is a constant. However, it is believed that equations (9) and (10) are valid, at least qualitatively, for magnetizations less than saturation, in which case the permeability curve for low fields goes to 1.0 as shown. Although ferrites may be operated near resonance for some isolator applications where large attenuations are desired, for other applications where small transmission losses are desired it is preferable to operate at fields as far below resonance as possible.

The permeability of Fig. 6 is the rf permeability which appears in Maxwell's equations, and is the ratio of internal b to internal h. It is independent of the shape of the ferrite body. However, the internal h may be quite different from the externally applied h. Since we are frequently more interested in the external h than the internal h, it is convenient to define the ratio of b internal to h external as an "external μ" (μ_{ex}). This external μ will go through resonance not at an internal field given by $\omega/|\gamma|$ but at a field which depends on the shape of the ferrite body. Thus, if the demagnetizing factors in directions normal to H_i are appreciable, a displacement of the total angular momentum vector from its equilibrium position will create restoring torques due to the creation of magnetic poles on the surface of the body. These restoring torques added to the torque produced by H_i will increase the resonance frequency of the body as a whole above $\omega_0 = -\gamma H_i$. Furthermore, anisotropy fields may also change the resonant frequency. It is therefore convenient to postulate a fictitious dc magnetic field "H effective" which takes into account H_i, the rf demagnetizing factors, and the anisotropy fields, and which if operating on an electron in free space would produce the same resonance frequency as is observed for the ferrite body.[7] By reading the abscissa of Fig. 6 as "H effective" and the ordinate as μ_{ex}, Fig. 6 may conveniently be used as a universal curve applicable for any shape of body for which an "H effective" can be defined. Since losses are a maximum at the resonance for μ_{ex} (not at the resonance for μ), the literature usually implies this μ_{ex} resonance when discussing ferromagnetic resonance. In this paper the unqualified term "resonant frequency" refers to the resonant frequency of μ_{ex}.

3. EXPLANATION OF FARADAY ROTATION IN TERMS OF NORMAL MODES OR
   COUPLED MODES

With a knowledge of the microwave properties of ferrites one can analyze the electromagnetic transmission properties of a ferrite-loaded medium but, in general, such an analysis is quite complex. However, the effect of ferrite loading of a two-mode transmission medium can be understood with relative simplicity using either of the following two viewpoints:

(1) the normal mode approach in which the input and output waves are resolved into two modes which are orthogonal in the presence of the ferrite,

(2) the coupled-mode approach, in which the behavior of the ferrite-loaded region is described in terms of a coupling, introduced by the ferrite, between two modes which are orthogonal in the absence of the biased ferrite.

In the following section both viewpoints are applied to an explanation of Faraday rotation in ferrite-loaded round waveguides.

3-1. *Normal Modes in a Ferrite Medium*

As Polder has pointed out, the fact that the permeability is a scalar quantity for circularly polarized fields means that in an infinite ferrite medium the normal modes of propagation along the Z-axis are $(+)$ and $(-)$ circularly polarized waves. In fact for considerably more complicated cases where, for example, the medium is a round rod of ferrite along the Z-axis in free space propagating dominant guided waves along the rod, it will still be true that circularly polarized waves will be the normal modes for the medium.

For a waveguide comprising a metal tube of circular cross section containing a very slender pencil of ferrite lying along the axis of the tube and having a diameter much less than the tube diameter, we can ascribe to the composite medium inside the metal tube an effective dielectric constant (ϵ_e) and effective permeability (μ_e) which for a homogeneous dielectric across the entire cross section would yield the same characteristic impedance and phase constant as is actually observed. Since the ferrite pencil was said to be slender, ϵ_e would be only slightly greater than one. Since the ferrite will perturb the dominant wave only slightly, circularly polarized dominant waves will produce circularly polarized transverse magnetic fields in the rod. The effective permeability for the waveguide would then be slightly greater or less than one for the $(-)$ and $(+)$ dominant modes respectively. Thus, the two circularly

polarized dominant waves which are the "normal modes" for the medium will travel at slightly different velocities. Alternatively, we can say that the index of refraction of the dielectric medium for $(+)$ waves is different from that for $(-)$ waves. In the usual case where the wave frequency is high and the H_i is low so that $\omega_0 < \omega$, the index of refraction for $(+)$ waves will be lower than the index for $(-)$ waves. Therefore, the $(+)$ waves will travel through the tube at a higher phase velocity than $(-)$ waves. As a result, a linearly polarized input wave traveling along H_i may be decomposed into $(+)$ and $(-)$ components which, after passing through the section, may be added to give a new linearly polarized wave whose polarization axis has been rotated in a clockwise sense relative to that of the input. The magnitude of the rotation depends upon the length of the section and also upon the effective dielectric constants and permeabilities. The difference in phase $\Delta\varphi$ between the $(+)$ and $(-)$ components passing through a length z of ferrite loaded waveguide is

$$\Delta\varphi = \ell(\beta_- - \beta_+) = \omega\ell\sqrt{\epsilon_e}(\Phi_-\sqrt{\mu_{e-}} - \Phi_+\sqrt{\mu_{e+}}) \qquad (11)$$

where Φ_+ and Φ_- are cut-off factors for the two polarizations which are determined by the geometry of the conducting boundaries and the μ_e and ϵ_e of the dielectric medium. If the operating frequency is well above cut-off, then to a first approximation Φ_+ and Φ_- equal one.

The angle of rotation for a linearly polarized wave will be

$$\theta = \frac{1}{2}\Delta\varphi = \frac{\omega\ell}{2}\sqrt{\epsilon_e}\ (\Phi_-\sqrt{\mu_{e-}} - \Phi_+\sqrt{\mu_{e+}}) \qquad (12)$$

We see that the Faraday rotation angle is dependent not only upon the difference between the square roots of μ_- and μ_+ but also upon the effective dielectric constant of the medium. Thus, if the medium is loaded dielectrically without altering the effective permeability, the Faraday rotation should increase.

3-2. *Coupled Modes in a Ferrite Medium*

What has just been outlined is the normal mode approach based upon the fact that the ferrite permeability is scalar for the two circularly polarized modes. Alternatively, we can use the coupled mode approach based on the use of linearly polarized modes and the associated tensor components of permeability. We recognize, as stated earlier, that the k part of the permeability tensor is a measure of the ability of the electrons to couple energy from a field of one polarization to a field of orthogonal

polarization. We know from the theory of distributed wave coupling[8] that if we have two modes of propagation in a waveguide having the same phase velocity, power can be completely transferred from one mode to the other if there exists uniform distributed coupling between the modes along the waveguide. In particular, if all of the energy is introduced as a wave in one of the modes, it will diminish in amplitude as it travels along the waveguide, while a wave in the other mode will start from zero and grow in amplitude. After a certain distance all of the power will have been transferred from the first to the second mode. Beyond this point the reverse process will take place. The amplitudes (voltages) of the two waves as a function of distance z along the waveguide will be given by

$$V_1 = V_0 \cos (KZ) \; \epsilon^{-j(\beta+K)Z+j\omega t} \tag{13}$$

$$V_2 = -jV_0 \sin (KZ) \; \epsilon^{-j(\beta+K)Z+j\omega t} \tag{14}$$

where K is the rate of change of voltage with distance of one wave produced by unit voltage in the other wave. Thus, the larger the coupling coefficient K, the more rapidly with distance will power cycle back and forth between the two modes.

Let us now apply this point of view to the round waveguide containing an axially magnetized pencil of ferrite, located along the waveguide axis. We may identify vertically and horizontally polarized dominant modes as the coupled modes for the medium. They do have the same phase velocity and they are orthogonal. When the ferrite pencil is magnetized, the spinning electrons produce a uniformly distributed coupling between the modes. If we launch a vertically polarized wave at the input it will decrease in amplitude as it travels along the waveguide and the horizontally polarized wave will increase in amplitude.

Let us assume that the X axis is vertical and that H_i and the direction of propagation are in the plus Z direction. In a thin transverse section of the waveguide, the h_y of the vertically polarized* wave induces through the spin-precession coupling an rf magnetization m_x which, for the usual case where $\omega_0 < \omega$, lags h_y by 90 degrees. This m_x then reradiates a horizontally polarized wave having an h_x which lags m_x by 90°.[9] Thus, h_x lags h_y by 180°, and the induced horizontally polarized wave will always be 180° out of phase with the vertically polarized wave (or in phase if $\omega_0 > \omega$). The resultant wave at any cross section will be linearly polarized and the angle of polarization will increase in a clockwise sense with distance in the direction of propagation. This is, of course, the same result

---

* As is customary, the direction of the electric field components is taken as the direction of polarization.

as was obtained using the normal mode approach. However, the coupled-wave point of view is very convenient for some problems. In particular, if we consider what will happen if our round waveguide is slightly flattened vertically, we can see immediately that the phase velocities for horizontal and vertical polarization will no longer be equal. Therefore, it will no longer be possible to get complete transfer of power from the vertically to the horizontally polarized modes. At the cross section where maximum power has been transferred to the horizontally polarized mode, the total wave will be elliptically polarized. At twice this distance the resultant wave will again be a pure vertically polarized wave, and the process will repeat cyclically along the axis of propagation.

The coupling coefficient K which is important to the coupled mode argument is obviously related to the effective permeability μ_e used in the normal mode argument, and either one should be derivable from the other. In deriving K directly we must first know the total transverse magnetic field (h_{xy}) at every point in the cross section for one of the coupled modes; and although we frequently assume an unperturbed dominant mode pattern for simplicity, strictly speaking we should use the pattern for the perturbed wave. Knowing the tensor permeability of the ferrite, we can deduce what the transverse b_{xy} at every point in the ferrite will be at a time one quarter period later than the applied field maximum. This induced b_{xy} will be at right angles to the h_{xy} which produced it and hence will couple to a new wave polarized at right angles to the original. However, in order to know how effective this b_{xy} is in generating the new wave, we must know how it conforms in relative magnitude and direction to the h for the pattern of the new wave. Since the coupling of the induced b to the new wave at each point is proportional to $(\vec{m} \times k\vec{h}_1) \cdot \vec{h}_2$, the total coupling coefficient is proportional to the integral of this product over the ferrite cross section.

$$K \approx k \int_A (\vec{m} \times \vec{h}_1) \cdot \vec{h}_2 \, da, \tag{15}$$

where $\vec{m}$ is a unit vector in the direction of the dc magnetization here assumed along the waveguide axis, $\vec{h}_1$ is the rf magnetic intensity produced in the ferrite by the original wave normalized to unit power, and $\vec{h}_2$ is the rf magnetic intensity which would be produced at the same point by the new wave normalized in the same fashion.

For points in the ferrite near the axis of a round waveguide, $\vec{m} \times \vec{h}_1$ for a linearly polarized TE_{11} wave is nearly conformal* with h_2 for the

---

* As used here, "conformal" does not relate to "conformal transformations," but is a measure of the similarity (in amplitude and direction) of two vector fields.

perpendicular polarization of the TE_{11} wave, and this region is very effective in producing coupling between linearly polarized waves. However for points in the ferrite far from the wave guide axis, the induced field will not be conformal with h_2, and consequently these regions of the ferrite will be less effective in producing coupling. Finally, since the Faraday rotation per unit length is equal to K, which is in turn proportional to the magnitudes of $\vec{h}_1$ and $\vec{h}_2$, it may be seen that anything which increases $|\vec{h}_1|$ and $|\vec{h}_2|$ for unit wave power will increase the Faraday rotation.

This procedure of integrating the product of the induced b with the h of the new wave field pattern is important in the determination of effective coupling between different modes in other instances as will appear later.

4. PRINCIPLES OF FERRITE BEHAVIOR IN FARADAY-ROTATION ELEMENTS
4-1 *Plane-Wave versus Loaded-Waveguide Theory*

The first microwave applications of ferrites made use of a round waveguide containing an axial ferrite pencil in an axial magnetic field. Such "Faraday plates" continue to be of great practical importance. We shall, therefore, endeavor to explain, by means of point-field concepts, the various considerations entering into the design of Faraday plates and to explain the peculiar behavior which they sometimes exhibit.

Let us first consider the case of a uniform plane TEM wave in an unbounded ferrite medium. With magnetization in the direction of propagation, the Faraday-rotation angle is related to distance in the direction of propagation by equation (12) provided we let the cut-off factors Φ_- and Φ_+ equal 1. The effective μ and ϵ will in this case be equal to the μ and ϵ for the ferrite itself. Using equations (9) and (10) for μ_+ and μ_- and assuming that

$$\omega \gg \omega_0 \quad \text{and} \quad \omega \gg \frac{\gamma M}{\mu_0}$$

equation (12) reduces to the following:

$$\theta = \frac{\gamma M \ell}{2} \sqrt{\frac{\epsilon}{\mu_0}} \tag{16}$$

This dependence of Faraday rotation on saturation magnetization and dielectric constant can be predicted qualitatively by making use of coupled wave theory. The Faraday rotation of a linearly polarized wave can be described as being due to a distributed coupling between the

vertically and horizontally polarized waves. This coupling is produced by the elementary magnetic dipole moments (electron spins) which precess about the direction of dc magnetization. The greater the magnetization, M, the greater the number of dipole moments per unit length of ferrite, thus causing an increase in coupling coefficient and in Faraday rotation. Similarly, if the dielectric constant of the ferrite is increased, the characteristic impedance will decrease, resulting in an increase in the magnitude of the rf h vector in the ferrite, by a factor of $\epsilon^{1/4}$, for a given amount of power in the wave. Since in the plane wave case $|\vec{h}_1|$ is everywhere equal to $|\vec{h}_2|$, the Faraday rotation is proportional to $\vec{h}^2$ (equation 15) and thus is proportional to $\sqrt{\epsilon}$.

Since the angle of rotation is proportional to the dc magnetization of the ferrite, one would expect a curve of rotation vs applied dc field to be similar to the saturation curve of the material. Furthermore, a curve of loss for circularly polarized waves as a function of applied dc field should reveal a high loss only near ferromagnetic resonance for the positive circularly polarized wave, with little loss at all applied fields for the negative wave. The plane wave theory also predicts that the angle of rotation should be independent of frequency so long as the frequency of the wave is much greater than the ferromagnetic resonance frequency.

Although the above picture is simple and explains the basic facts of Faraday rotation, it is quite inadequate for the ferrite-loaded waveguide problem. To choose a particular example, the dominant TE_{11} wave in a circular waveguide has an H pattern which is transverse only at the center, and may be either longitudinal or transverse near the walls of the waveguide, depending on azimuthal location. These different parts of the magnetic field pattern react with any ferrite present in quite different ways, so that the situation becomes much more complicated than for the simple plane wave case; in the latter condition h is everywhere transverse and reacts with the ferrite in the same way at every point These complications may cause various apparently "anomalous" effects which can readily be explained in a qualitative manner by examining the rf magnetic field pattern microscopically and determining the interaction of the wave and the ferrite at each point. We shall therefore discuss some of the problems involved in designing Faraday rotation elements in circular waveguides, emphasizing the above point of view.

4-2. *The Impedance-Match Problem*

One of the first problems confronting the designer of Faraday rotation elements is that of obtaining an impedance match into the ferrite

whose dielectric constant is typically in the range from 10 to 20. If no attempt is made to obtain an impedance match, reflections from the ends of the sample of ferrite will occur and will interfere with each other, depending upon the electrical length of the sample. Since the effective permeability of the ferrite is a function of the applied magnetic field, the electrical length of the sample is a function of magnetic field. As a result, end reflections can and do cause serious aberrations or "anomalies" in the observed Faraday rotation and transmission loss as the magnetic field is varied.

The use of quarter-wave impedance-matching transformers or other such simple lumped structures is not an easy solution to the ferrite matching problem for the simple reason that characteristic impedance and the phase velocity for the ferrite will depend upon the magnetic field applied and also upon the direction of rotation of circularly polarized waves. A simple matching transformer can produce a match for the dielectric discontinuity at the ferrite ends, and in cases where this is the dominating effect the quarter-wave transformer may be attractive. However, when the magnetic effects are included there is no length of ferrite-loaded line which is truly a quarter-wave long for both the positive and negative circularly polarized waves which compose any linearly polarized wave, and when this difference in phase constants is large the quarter-wave transformer is not attractive. Continuous impedance tapering, however, appears to be a method which will always work for both directions of polarization as well as for all values of magnetic field. Thus, a ferrite cylinder may have its ends ground down to conical points, or it may be diluted at the ends by mixing with some other dielectric so that the effective dielectric constant tapers to a low value at the ends of the sample. We have found that mechanically tapering the ends of the sample is an effective means of producing impedance match provided that the diameter of the waveguide and of the sample is such as to permit only the dominant wave to be propagated in the ferrite-loaded section.

4-3. *Limitation on Ferrite Diameter*

The relatively high index of refraction of most ferrites also introduces a mode conversion problem in the design of Faraday rotators. Thus, a dominant-mode hollow metal waveguide may, when fully or even partially loaded with a ferrite rod, be capable of propagating a higher order wave. If the ferrite is tapered and placed along the axis of a waveguide propagating the dominant TE_{11} mode, there will in fact be a tendency to convert to the TM_{11} mode, as has been demonstrated experimentally

for pencils of high dielectric constant. This mode conversion tendency can be understood qualitatively by referring to the magnetic field patterns of the TE_{11} and TM_{11} modes as shown in Fig. 7. For the TM_{11} mode the h at the center is 180° out of phase with the h at the periphery, whereas for the TE_{11} mode the center and peripheral h lines are in phase. By delaying the center h lines of the TE_{11} mode relative to the h lines near the periphery, the tapered pencil would tend to convert the TE_{11} mode to the TM_{11} mode. From symmetry considerations one can show that the TM_{01}, TE_{21} or TE_{01} modes which have cut-offs between the TE_{11} and TM_{11} modes will not be generated unless the ferrite is inhomogeneous.

This mode conversion problem becomes evident when one makes measurements of rotation, loss, and ellipticity on a large diameter, properly tapered ferrite rod, as a function of magnetic field. Instead of the smooth variations expected from plane wave theory, peculiar large variations of the above quantities are observed. We now believe that these variations can be attributed to a partial conversion of the dominant TE_{11} mode to the TM_{11} mode by the ferrite pencil which is of sufficient diameter to permit the propagation of both modes. As the magnetic field is varied, the effective electrical length of the ferrite changes so that the interference between the TM_{11} and the TE_{11} modes can also vary, thereby causing fluctuations in transmission and rotation. Furthermore, such a sample will show relatively large reflections, of the order of 10 db below the incident wave, notwithstanding the fact that it may be equipped with slender tapers of sufficient length that good impedance
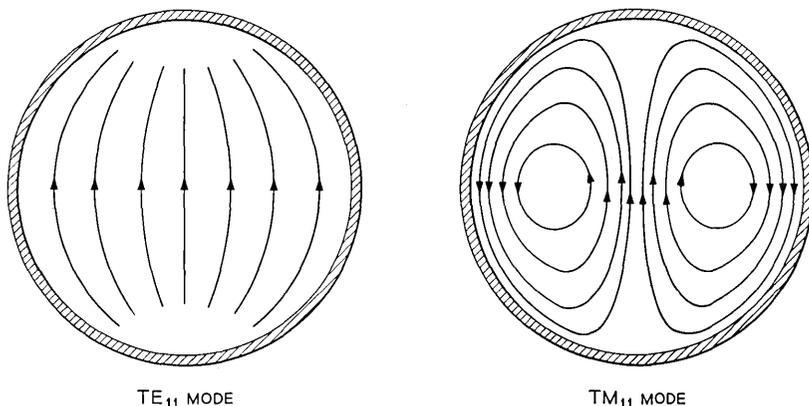
TE$_{11}$ MODE   TM$_{11}$ MODE

Fig. 7 — Patterns of magnetic intensity for the TE_{11}O and TM_{11}O Modes.

match would be expected. This also is to be explained in terms of the presence of a TM_{11} wave within the sample which may be multiply reflected internally and which can give rise to reflected TE_{11} power by reconversion at the input end.

The above mode conversion hypothesis has been checked in the following manner. Let us consider the TE_{11} mode as being composed of two circularly polarized TE_{11} waves rotating in opposite directions. Since the theoretical permeability for the negatively rotating component is larger than the theoretical permeability for the positively rotating component
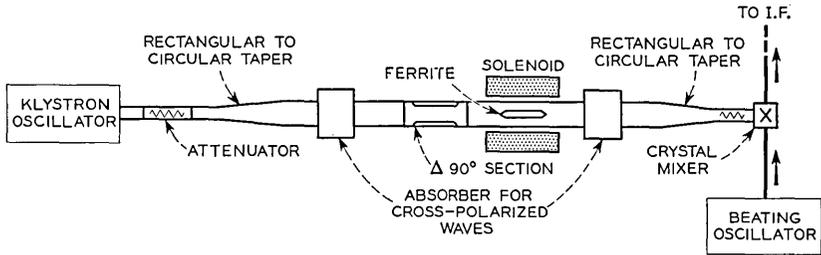
Fig. 8 — Apparatus for the measurement of ferrite-loaded round guide with circularly polarized $TE_{11}O$ waves.
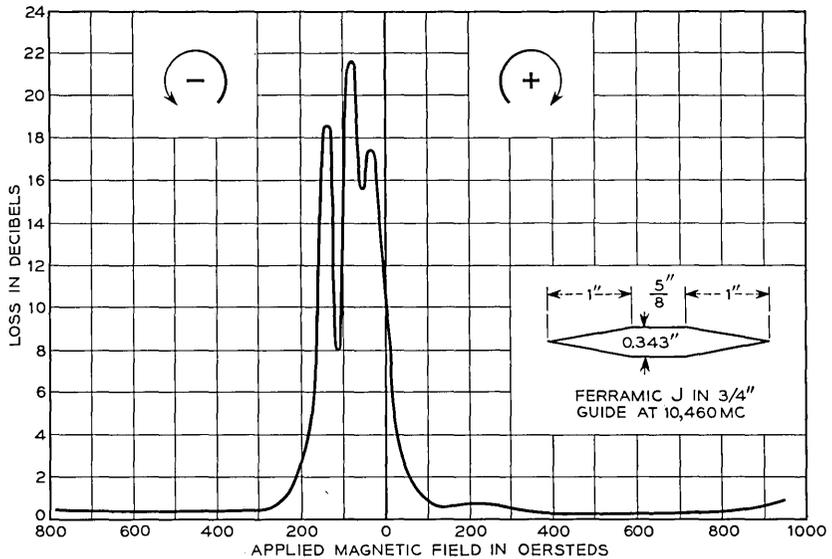
Fig. 9 — Measured curve showing mode conversion effects in ferrite-loaded round guide at low longitudinal fields for negative circularly polarized waves.
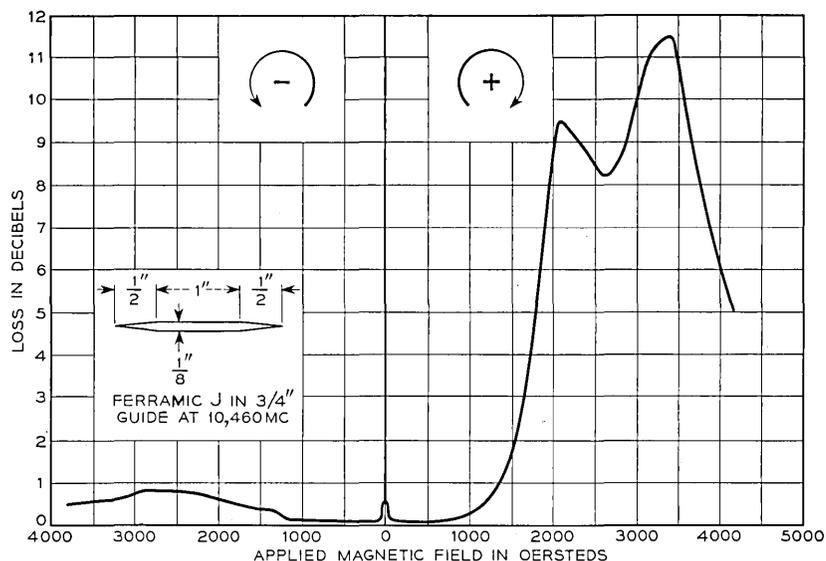
Fig. 10 — Curve showing ferromagnetic resonance in ferrite-loaded round guide.

(see Fig. 6), we might expect that the negatively rotating component would be more readily subject to mode conversion than the positively rotating component. This supposition has been verified experimentally. Circularly polarized waves were generated by means of a 90° "squash" section and were transmitted through samples of ferrite as shown by the circuit of Fig. 8. The resulting curves of transmission loss vs magnetic field showed that for ferrite samples of large diameter there were many violent fluctuations of loss for both positive and negative circularly polarized waves. For a somewhat smaller diameter, fluctuations in loss for positive waves diminished substantially or disappeared entirely while the fluctuations for negative waves remained, as shown in Fig. 9. For still smaller diameters the fine-grain fluctuations disappeared entirely in the zero to 500 oersted region for both positive and negative circularly polarized waves, as shown in Fig. 10. This, we feel, confirms the hypothesis that the loss and rotation fluctuations are caused by mode conversion. Therefore, the diameter of ferrite used in Faraday-rotation devices must be small enough so that no mode conversion takes place.

Even if the dielectric constant of the ferrite were low enough so that no mode conversion problems would arise, one would still confine the ferrite to the axial region of the waveguide, since it is in this region that the ferrite is most effective as a rotator while the peripheral portions may

contribute to loss but not to rotation. There are various ways of making the above statement evident, but perhaps the easiest is to consider the rotation as being produced by the elementary electrons which induce an rf magnetic field perpendicular to the exciting rf field. Therefore, near the axis of the guide this induced field will set up a TE_{11} mode perpendicular to the input field and thus cause rotation. Near the periphery, however, the electrons will tend to set up rf magnetic fields normal to the wall of guide which cannot reradiate as a TE_{11} mode since no such field component exists for the TE_{11} mode. Thus, the peripheral portions of the ferrite contribute nothing to the coupling coefficient K given by equation (15). Therefore, the peripheral portions of the ferrite will not be effective as a Faraday rotator but may produce dielectric loss due to the peripheral electric field as well as magnetic loss due to the peripheral magnetic field.

4-4. *Dependence of Rotation and Loss on Geometry of Ferrite*

Realizing the necessity of confining the ferrite to the axial region in order to prevent multimoding and in order to have the ferrite in the most effective portion of the guide for rotation, one is still confronted by the problem of determining the diameter which will yield a maximum ratio of rotation to loss. This ratio may be used as a figure of merit, F, defined as the Faraday rotation in degrees per db transmission loss.

We shall first discuss the variation of rotation as a function of diameter of the ferrite. From the naive plane wave picture, one would expect the rotation to be directly proportional to the number of precessing mangetic moments, and thus to the mass of ferrite or the square of the ferrite diameter. For small diameter ferrite rods, the measurements recorded in Fig. 11 demonstrate that rotation per gram is indeed independent of diameter, for small diameters.

Theoretical guidance on this same problem can be found in the work of L. R. Walker and H. Suhl,[10] who have solved mathematically for the magnitude of Faraday rotation to be obtained from thin pencils of ferrite in a circular waveguide. This mathematical treatment also predicts that for small diameter pencils the rotation should be proportional to the square of the diameter of the ferrite.

For pencils of large diameter, no general mathematical treatment which takes into account the dielectric constant of the ferrite is available. However, for the particular case of a ferrite of dielectric constant 10 (typical for most ferrites) inside a circular waveguide with a diameter to free-space-wavelength ratio of 0.8, Suhl and Walker as well as H.

Seidel[11] of Bell Telephone Laboratories have obtained values of loss and rotation as a function of ferrite diameter. Fig. 12 is a theoretical curve of rotation per mass as a function of ferrite diameter based on calculations of Suhl and Walker, while Fig. 11 shows a typical experimentally obtained curve. For very small ferrite diameters, the rotation per mass is seen to be independent of ferrite diameter as explained previously. With further increase in ferrite diameter, both curves rise sharply, while the theoretical curve reaches a peak and then declines. With most typical ferrites one cannot usually observe this peak because multimoding problems become serious before this large diameter is reached. However, with some ferrites of lower dielectric constant, such rotation per gram peaks have been observed.

The above results can be explained in the following manner. Any increase in diameter of a dielectric rod above a certain minimum value increases the energy concentration in the dielectric much more rapidly than the increase in cross-sectional area of the dielectric. This has been shown to be the case by H. Seidel,[11] and also seems reasonable by analogy from the dielectric waveguide studies of M. C. Gray.[12] Thus, as the
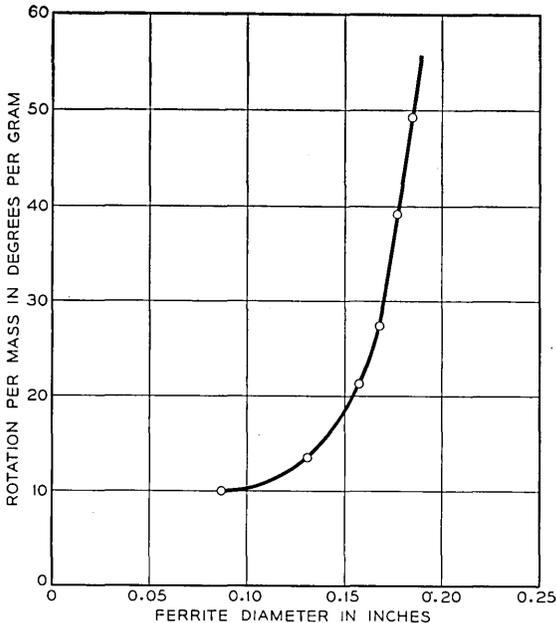
Fig. 11 — Measured rotation per ferrite mass versus ferrite diameter for 0.75″ I.D. guide at 11.2 kmc.
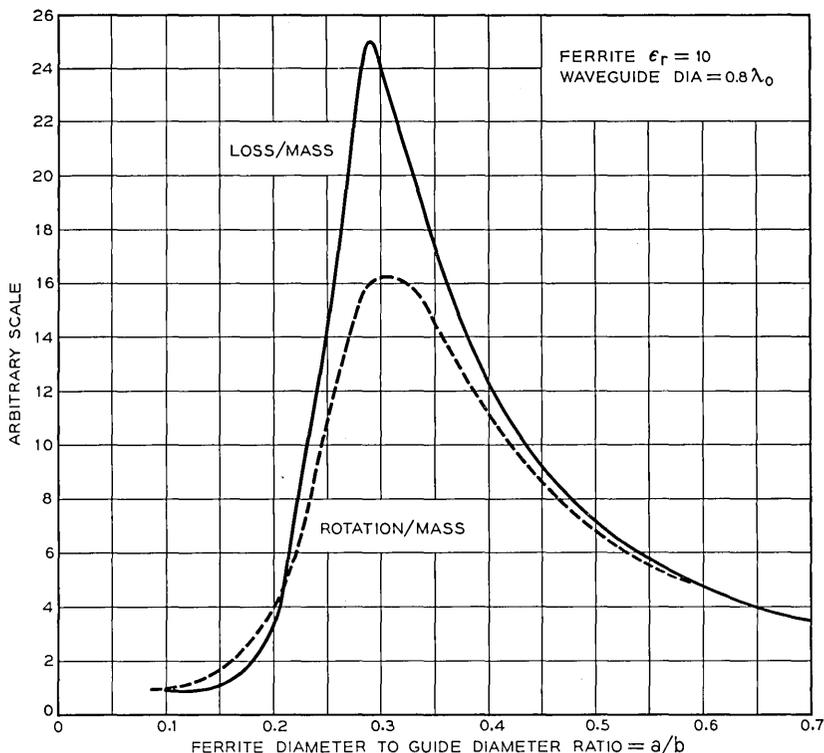
Fig. 12 — Calculated rotation and dielectric loss per ferrite mass versus ferrite diameter.

diameter of the ferrite is increased, the energy concentration in the ferrite increases rapidly, first for the negative circularly polarized wave whose permeability is above unity, and then for the positive wave whose permeability is below unity. Therefore, the rotation per gram as well as the rotation will increase with ferrite diameter in this range of diameters.

Another consequence of an increase of ferrite diameter is an increase in the effective dielectric constant of the composite air-ferrite guide, thus decreasing the characteristic impedance of the guide and thus resulting in an increase in the magnitude of the transverse rf h vector in the ferrite for a given amount of input power. This too can result in an increase of Faraday rotation per gram with increasing ferrite diameter entirely apart from any redistribution of energy in the waveguide's cross section.

The combination of the dielectric waveguide effect and the lowering

of the effective characteristic impedance results in a broad maximum value of rotation which occurs at a diameter for which the difference in the propagation constants of the positive and negative rotating components is a maximum. Increasing the diameter in this range will increase the mass of the ferrite without appreciably increasing the rotation, thus decreasing the rotation per mass.

We shall now discuss the variation of loss with diameter so as to be able to determine the figure of merit, F, (rotation/db loss) as a function of diameter. The loss in the ferrite can be due to dielectric loss or to magnetic loss. Let us first consider a ferrite for which the loss is entirely magnetic. This loss will be proportional to the square of the transverse magnetic field since the longitudinal rf magnetic field will cause no losses if the ferrite is saturated in the longitudinal direction.

Except for large ferrite diameters both the loss per mass and the rotation per mass are proportional to the square of the transverse magnetic field so that the figure of merit is independent of ferrite diameter regardless of how transverse h varies with diameter. This constancy of the figure of merit for ferrites with magnetic losses only is confirmed by the Suhl and Walker analysis.*

For large diameter ferrites, some of the transverse h does not contribute to rotation (as discussed in Section 4.3) while all of the transverse h contributes to loss. Therefore, one would expect the figure of merit to decrease with increasing diameter for large-diameter ferrites having magnetic losses. However, it appears from the Suhl and Walker analysis that this effect is small until the waveguide is almost completely filled.

The above discussion assumes that the operating frequency is far above the resonance frequency. The situation is more complicated at low microwave frequencies where, due to proximity to ferromagnetic resonance, the loss for the positive circularly polarized component may be higher than for the negative component.

Let us next consider the case of a ferrite having dielectric loss and no magnetic loss. H. Seidel[11] has made theoretical computations of dielectric loss as a function of ferrite diameter with zero applied magnetic field. This data has been plotted as loss per mass as a function of ferrite diameter in Fig. 12, and is to be compared to the rotation per mass curve of the same figure. It is seen that the loss per mass curve rises sharply with

---

* See Reference 10, p. 1152. Using their notation, rotation is proportional to $\kappa'Q$ while loss is proportional to $\mu''P + \kappa''Q$. Far from resonance $\kappa'' \ll \mu''$ so that the loss is proportional to $\mu''P$ and the figure of merit is proportional to $\kappa'Q/\mu''P$. But from Fig. 3 in the Suhl and Walker paper, P/Q is approximately unity for most values of ferrite diameter. Therefore, the figure of merit will also remain approximately constant over a wide ferrite diameter range.

increasing ferrite diameter, reaches a peak near $a/b = 0.3$ and then declines. The first sharp rise is due to two causes: first, the increase in energy concentration in the ferrite because of the dielectric waveguide effect discussed previously in connection with the rotation per mass curve; and second, because of the sharp increase in the longitudinal electric field in such a shielded dielectric waveguide. As the ferrite diameter increases beyond the point where most of the electromagnetic energy is in the ferrite, the loss due to the transverse electric field does not increase, while the longitudinal electric field decreases until the latter is zero in a completely filled waveguide. Therefore, the loss will drop with increasing ferrite diameters in the large diameter region, resulting in an even sharper drop in loss per mass in this region.

The figure of merit as a function of ferrite diameter will be given by the ratio of the rotation to the sum of the magnetic and dielectric losses. Since the figure of merit due to magnetic loss is essentially independent of diameter, the variation in the figure of merit with diameter depends on the ratio of rotation to dielectric loss. One is usually interested in ferrite-to-waveguide diameter ratios below about 0.3. In this region both the rotation and dielectric-loss curves rise sharply (Fig. 12), but their ratio first rises to a peak at a ferrite-to-waveguide diameter ratio of about 0.125 for the typical conditions shown in Fig. 12. At larger ferrite diameters the theoretical figure-of-merit drops and then rises again at ferrite-to-waveguide diameter ratios exceeding 0.275. Fig. 13 shows experimental curves obtained by J. P. Schafer for the figure of merit as a function of ferrite diameter for two typical low-loss ferrites. These curves have a shape which agrees with theoretical expectations, and confirm that there is an optimum diameter for maximum figure of merit.*

4-5. *Rotation as a Function of Frequency*

The plane-wave analysis (Section 4-1) shows that the rotation is independent of frequency if the operating frequency is much higher than the resonance frequency. In a waveguide, however, the rotation will increase with increasing frequency for two reasons. First, since the impedance of a TE mode in a waveguide decreases with increasing frequency, the transverse rf magnetic field increases as the frequency gets

---

* It must be emphasized that the above discussion assumes that the ferrite sample is long compared to its diameter so that changes in diameter will not seriously affect the demagnetizing factors and the uniformity of field in the ferrite. If, with constant applied field, the diameter of a ferrite cylinder is increased so that its length-to-diameter ratio is no longer large, the increase in demagnetizing factor may cause a loss in saturation, thus increasing the low field losses and decreasing the figure of merit.
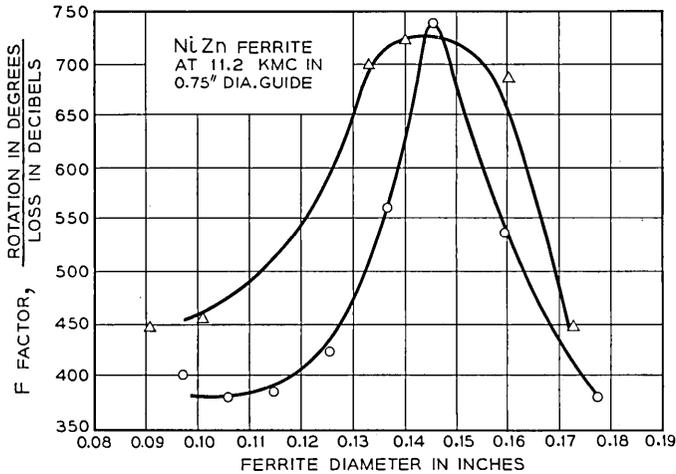
Fig. 13 — Faraday rotation per db loss as a function of ferrite diameter.

farther above cut-off, thus resulting in an increase in rotation. Second, for a waveguide partially loaded with ferrite, an increase in frequency causes the electromagnetic energy to become concentrated more and more in the ferrite rod, which acts as a dielectric waveguide as described in Section 4-4. Since for dielectric waveguides it is the ratio of diameter to wavelength which determines the ratio of energy in the ferrite to the total energy in the guide,[12] a decrease in wavelength should have the same effect on the energy concentration in the ferrite as an increase in diameter. Therefore, for diameters in the steep part of the curve of rotation per gram versus diameter, (for example, Fig. 12), one should expect a rapid increase in rotation with increasing frequency. For large ferrite diameters for which all of the electromagnetic energy is in the ferrite (or for high enough frequencies) the rotation should be independent of a further increase in frequency. Fig. 14 shows an experimental curve taken by J. P. Schafer of the ratio of the rotation at 11.7 kmc to the rotation at 10.7 kmc as a function of diameter in the small diameter range. This curve is in agreement with the above discussion.

4-6. *Dependence of Rotation on Dielectric Constant of Surrounding Medium*

J. P. Schafer and J. H. Rowen[13] have observed an enhancement of rotation for a partially ferrite-loaded guide with a high dielectric constant surrounding medium over that observed with a polyfoam (ϵ = 1.05) surrounding medium. Our own observations have shown both an

increase and a decrease of rotation with an increase in the dielectric constant of the surrounding medium, depending upon the diameter of the ferrite. Embedding a small-diameter ferrite pencil in a high dielectric constant support increases the effective dielectric constant of the guide, thus increasing the transverse magnetic field in the ferrite and increasing the rotation. If the loss in the ferrite is primarily dielectric in nature, the figure of merit will also improve because of the decrease in the electric field in the ferrite.

For ferrites of larger diameter, however, the dielectric waveguide effect becomes important and the ratio of energy in the ferrite to the total guide energy decreases with increasing dielectric constant of the surrounding medium. This results in a decrease in the rotation rather than the increase noted by Schafer and Rowen.

4-7. *Dependence of Rotation and Loss on Temperature*

As shown by equation (16), the rotation is directly proportional to the saturation magnetization, M. It is well known, however, that M decreases with increasing temperature, going to zero at the Curie temperature.[14] Therefore Faraday rotation will also decrease with increasing
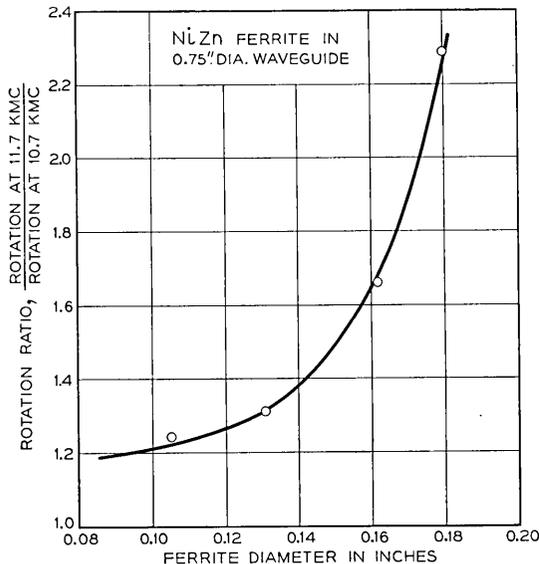
Fig. 14 — Frequency sensitivity of Faraday rotation as a function of ferrite diameter.

temperature. At 11 kmc for a ferrite with a Curie temperature of 150° C, we have measured a variation of 0.4 per cent rotation per °C at room temperature. Ferrites having a much higher Curie temperature show a much smaller variation of rotation with temperature. In fact, measurements have been made with ferrites having a Curie temperature of 500° C, which showed less than 0.03 per cent change in rotation with variations of temperature of 20° C near room temperature.

Magnetic losses are also affected considerably by temperature. It is known* that anisotropy in ferrites drops rapidly as one approaches the Curie temperature. If the magnetic losses are due to anisotropy, then heating the ferrite should also reduce the losses. Experimental results on fairly low Curie temperature ferrites indicates that this is indeed the case. Thus, for a particular sample at 10,000 mc, the loss at saturation dropped from 3 db at room temperature to 1 db at 80° C. For ferrites having high Curie temperatures, the loss does not necessarily decrease with increase in temperature since the anisotropy variation at temperatures far below the Curie temperature may be rather complex.[15]

The variation of figure of merit with temperature depends on whether the loss drops more or less rapidly than the rotation drops with increasing temperature. The figure of merit of some ferrites has been observed to improve and others to drop with increasing temperature.

4-8. *Idiosyncracies of Ferrites Used as Faraday Rotation Elements*

   a. Loss Near Zero Field

It has been known for some time that some ferrites have rather large loss at low magnetic fields, and that with the application of moderate magnetic fields in any direction, longitudinal or transverse, this loss disappears. This low-field loss has been explained in a variety of ways but, in general, these all depend upon the presence within the ferrite of domain walls. Consequently, with fields strong enough to saturate the ferrite, the domain walls and the loss disappear. One would expect that if this loss were measured with circularly polarized waves, it should be independent of the direction of circular polarization used. Actually, however, our experiments on small-diameter pencils, which are free of mode conversion effects, have shown that this low-field absorption has a peak for counterclockwise polarized waves as seen in Fig. 15.

This puzzling asymmetry can be explained by the fact that more of the energy will travel within the ferrite for the negative circularly polar-

---

* C. Kittel, Reference 14, p. 182.
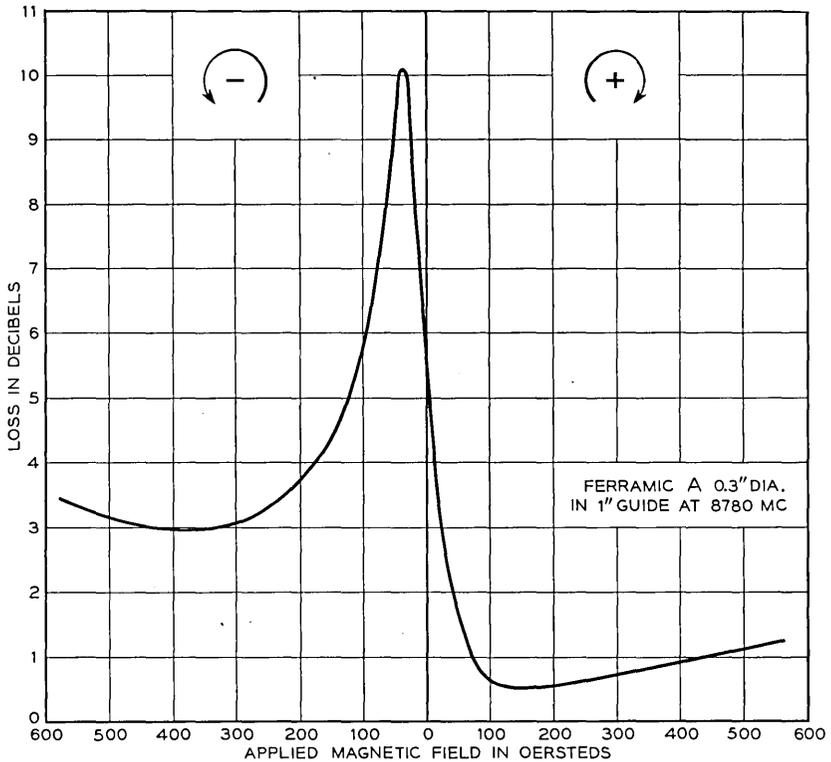
Fig. 15 — Measured curve showing displacement of low field loss peak.

ized wave than for the positive wave because of the higher index of re-
fraction of the former (see Fig. 6) as explained in Section 4-4. As a result,
the low-field loss can be expected to be greater for the negative wave as
observed experimentally.

b. Ferromagnetic Resonance for Negative Circularly Polarized Waves

Our measurements on slender pencils of ferrite using circularly polar-
ized waves disclosed another wholly unexpected result. One would expect
to observe a single loss peak, at fields high enough to produce ferromag-
netic resonance, for the positive circularly polarized wave. However, in
addition to this expected resonance, another resonance was observed for
negative circularly polarized waves at magnetic fields of nearly the
same value as that required to produce ferromagnetic resonance for the
positive wave as shown in Fig. 16. The magnitude of this high-field peak

for the negative wave depends upon the size of ferrite pencil used, being very small for very thin pencils (Fig. 10) and reaching values above the loss peak for the positive wave for some larger diameter samples (Fig. 16).

It is believed that this effect can be explained by analyzing the rf magnetic field at different points in the ferrite. Thus, by reference to a field pattern for the TE_{11} mode (Fig. 7), it can be seen that a rotation of this pattern as a whole, corresponding to a circularly polarized wave, causes the rf magnetic vectors near the center to rotate, indicating that they are circularly polarized. However, the magnetic vectors at the wall of the waveguide are everywhere tangent to the wall, and are linearly polarized even though the wave as a whole is circularly polarized. Since a linearly polarized vector consists of both positive and negative circular polarizations, any ferrite located off the center of the waveguide will see a certain amount of positive polarized field vector with its attendant loss even though the driving wave is negatively polarized. Consequently, one would expect to observe a small loss peak for a negative circularly

Fig. 16 — Measured curve showing ferromagnetic resonance for both positive and negative circularly polarized waves.

polarized wave, but this alone would not appear to account for the high loss peak obtained with the large diameter ferrites.

The high negative wave loss observed in the larger diameter samples can perhaps be attributed to TM_{11} mode propagation. By examining the transverse magnetic field pattern for this mode, it can be seen that when the rf magnetic field vector is circularly polarized in the negative sense in the center of the guide, it is polarized in the positive sense near the periphery. This is substantiated by the theoretical work of H. Suhl and L. R. Walker[16] which shows that the propagation constant of both positive and negative circularly polarized waves are nearly equal in a ferrite-filled circular waveguide propagating the TM_{11} mode. For this mode, the ferrite cannot distinguish between a positive or negative circularly polarized wave. Therefore, one would expect equal loss for the positive and negative waves at the resonance magnetic field value.

When one has a mixture of TE_{11} and TM_{11} waves in the ferrite, the field at the center may be stronger or weaker than the field near the periphery, depending upon the relative phase of the two waves. Thus, the negative-wave loss might be greater or less than the positive-wave loss, depending on phasing. It is indeed probable that the TM_{11} wave was present for the case of Fig. 16 since the ferrite was sufficiently large to permit anomalies in the low field region which are normally associated with multimoding, and the geometry is one which we know can generate the TM_{11} wave.

It may be noted that the ferromagnetic resonances shown in Fig. 16 have a double humped character. Similar tests made on other ferrites did not reveal any such double hump. In order to determine whether the double hump in Fig. 16 is a characteristic of the ferrite composition and structure, W. A. Yager measured the ferromagnetic resonance of a tiny sphere of this material at 24 kmc using his cavity technique. Yager did not find a double hump, indicating that the geometry and size may be responsible in our observations.

c. Molded Powdered Ferrites ·

Techniques have been developed for powdering ferrites and then molding them with polystyrene powder under heat and pressure. This process has the advantage of making it easy to obtain samples of any desired size with tapers molded in. Furthermore, it enables one to choose a wide variety of densities so as to obtain a variety of values for effective dielectric constant and for rotation per unit length.

However, measurements at 10 kmc show that the loss is much higher

for molded or powdered ferrites than it is for solid ferrites. This is not due to the molding process but can perhaps be explained on the basis of the broadening of resonance in the molded samples due to the wide variety of demagnetizing factors in the little powder particles. At 20 kmc and higher this increase in loss for molded samples should disappear since even a broadened resonance would not have much effect at the comparatively low fields required for saturation. Although this has not been checked in detail, in general it has been observed that molded samples of certain ferrites have extremely low loss at 24 kmc.

4-9. *Faraday Rotation Devices*

We shall now briefly describe some applications of Faraday rotation elements.

The gyrator function is performed by a Faraday rotation of 90° as described by C. L. Hogan.[4]

The Faraday-rotation type of circulator, due to S. E. Miller,[17] is illustrated in Fig. 17. Experimental models have been built for operation at 4, 11 and 24 kmc. The Faraday rotator for the 24-kmc circulator contains a $\frac{1}{8}''$ diameter tapered ferrite rod mounted in a polyfoam holder in a $\frac{3}{8}''$ round waveguide. The longitudinal magnetic field is supplied by a permanent magnet of such field strength that the ferrite produces 45° rotation. Sections 1 and 2 are rectangular-to-round transitions while sections 3 and 4 are cross-polarization pick-offs of a type designed by A. P. King of Bell Telephone Laboratories. The rectangular guide portions of these latter elements accept all of the wave energy in the round waveguide which is polarized orthogonal to the output in the rectangular-to-round transitions. Thus, an input at 1 is rotated by 45° in passing through the ferrite section and is transmitted to 2. Similarly, an input at 2 will go to 3, at 3 to 4, and at 4 will go back to 1 with a 180° phase reversal.

By terminating terminals 3 and 4 in the above circulator, one has an isolater which will transmit energy from 1 to 2, but any energy entering at 2 will be lost in the termination at 3. However, for this isolator application, there is no need for the cross polarization pickoffs since these can be replaced by diametral resistance-sheet terminations in the circular waveguides near terminals 1 and 2. These sheets are oriented in such a manner as to absorb all energy polarized perpendicular to terminals 1 and 2. Isolators of this kind have already proved very useful in preventing oscillator pulling when operating into mismatched loads.

In the above isolator, the ferrite element must be adjusted for exactly

45° of Faraday rotation in order to be theoretically capable of infinite attenuation in the reverse direction. Since the rotation angle is a function of frequency and temperature, one would expect the reverse attenuation to vary with frequency and temperature, and this will limit the bandwidth of the device.

The frequency and temperature range over which satisfactory isolation is provided can, however, be extended by operating several isolators in tandem as follows: Fig. 18 shows an isometrically exploded view of three isolators operated in tandem. Power is introduced at the left through a rectangular-to-circular waveguide transition and enters the system polarized as shown by c_1. After passing through the three Faraday rotators it emerges polarized as e_4, which is then accepted by a circular-to-rectangular transition. R_1, R_2, R_3 and R_4 are cross polarization absorbers designed to absorb reflected wave components. It can be seen that this assembly is in reality merely three isolators in tandem, where intervening transitions from circular-to-rectangular waveguide have been omitted as unnecessary. If all of the Faraday rotators are designed to produce exactly 45° of rotation at some one temperature, then the return loss for the whole system would be infinite at this temperature, even with a reflectively terminated receiving end. At some other temperature where the return loss for each isolator would be x db,
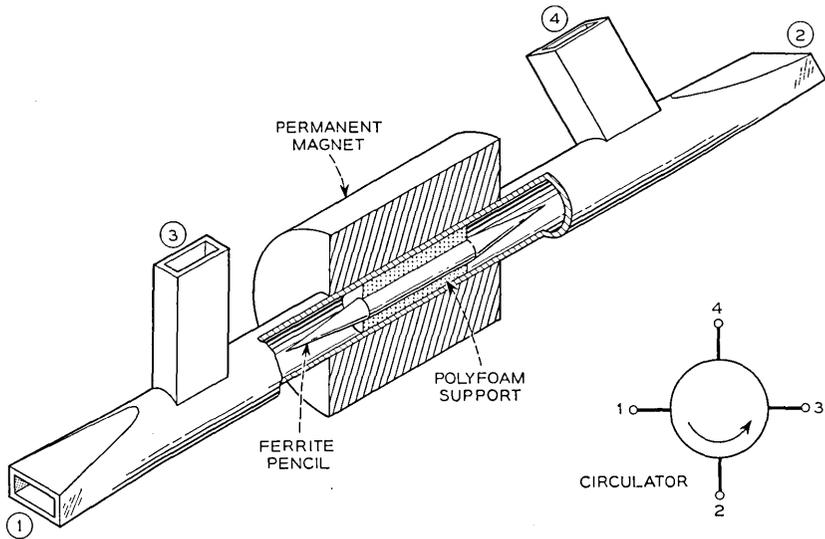
Fig. 17 — Faraday rotation circulator.

the total return loss for the system would now be increased to $3x$ db. Naturally, the forward transmission loss would also be increased by a factor of 3. If, however, this forward transmission loss is sufficiently small for a single isolator, an increase of transmission loss of three times may be a small price to pay for an increase in the return loss of three times. While tandem operation of three isolators is discussed here, any other number may of course be used. J. P. Schafer[18] has developed a two section isolator of this type for 11 kmc, and reports a reverse loss of 53 db and a forward loss of 0.25 db.

The above adjustment of the isolators gives a sort of maximally flat variation with temperature as indicated schematically in Fig. 19(a). A somewhat different type of broadbanding with temperature can be obtained by designing each of the isolators to produce exactly 45° of rotation at 3 different temperatures in the desired operating range. The return loss characteristic as a function of temperature will then look somewhat as shown in Figure 19(b).

The proposals made above for widening the operating temperature range should also be suitable for widening the frequency band of an isolator. In this case the tandem isolators of Fig. 18 might be designed to produce exactly 45° of rotation at the same frequency, or alternatively to produce 45° of rotation at three different frequencies in order to increase the return loss over the band.
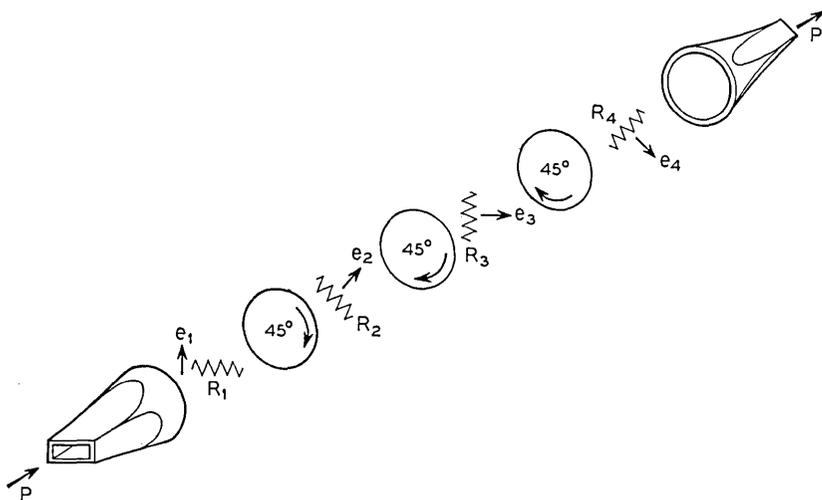
Fig. 18 — Schematic of a broadband isolator.
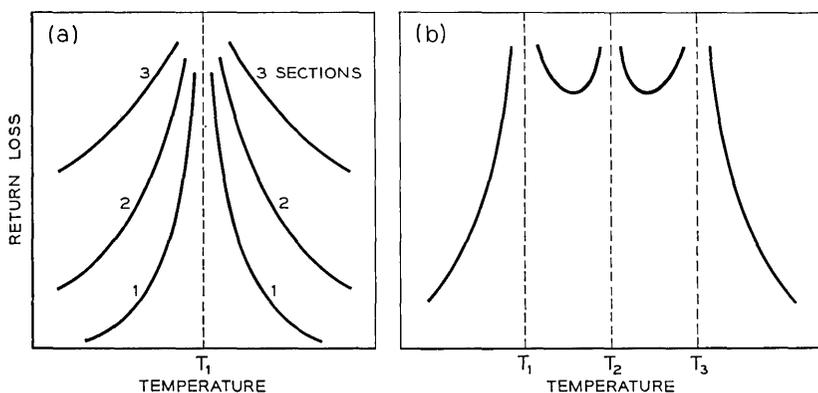
Fig. 19 — Alternative design characteristics for the isolator of Fig. 18.

It has been assumed that the cross-polarization absorbers R_1, R_2, R_3, R_4 in the compound isolator absorb 100 per cent of the power polarized parallel with the card. If they do not absorb all of the incident power, the infinities of the loss curves of Fig. 19 (a) and Fig. 19 (b) become finite. However, other infinities may exist due to cancellation effects. In particular, for the 2-section isolator of Fig. 19(a), a new infinity will appear on either side of the center of symmetry (T_1). The spacing of these infinities from the center will be proportional to the transmission past the absorbers. Schafer[18] has developed a two-section isolator employing broadbanding of this type to obtain a reverse loss in excess of 30 db over the band from 10.7 kmc to 11.7 kmc.

The above circuits utilize 45° of Faraday rotation. For other rotation angles, the devices may have a variety of peculiar properties. For example, if $22\frac{1}{2}°$ of Faraday rotation is used with the terminal polarizations shown in Fig. 20, the device becomes a non-reciprocal power divider with the properties indicated. Between certain pairs of terminals the power transmission is the same as in a circulator, while in the reverse direction there is a fifty-fifty power division as in a hybrid.

4-10. *Faraday Rotation in Non-Circular Guides and for Other Modes of of Propagation*

Faraday rotation depends upon the existence of two degenerate modes of propagation in the ferrite-loaded waveguide. For the discussion up to this point the two modes were the two polarizations of dominant $(TE_{11}°)$ waves, which are degenerate in the sense that they have identical phase and attenuation constants, and are orthogonal in the sense that

separate connections may be made at each end of the ferrite-loaded medium from the two TE_{11} waves to a pair of single-mode waveguides (illustrated in Fig. 17). It is important that the ferrite have a cross-sectional shape which is symmetrical and that the waveguide be circular in order that the two TE_{11}° waves have identical phase constants. If the guide is not round, if the ferrite is not symmetrical (or not homogeneous) in cross-section, or if the ferrite is located off the axis of the cylindrical guide, the phase constants of the two TE_{11} waves will be unequal; then for a linearly polarized input wave the output of the ferrite-loaded section will be elliptically polarized, and it will be impossible (in the absence of phase-correcting appendages) to achieve complete decoupling between the various terminals in an isolator or circulator.

The ferrite-loaded guide may be square instead of round, and for such a structure Faraday rotation of dominant-wave ($TE_{10}^{\square}$) polarization would take place. Applications directly analogous to those for round guides may be drawn.

Faraday rotation may be utilized with other pairs of degenerate modes. For example, in round guide the TE_{12}, $TE_{13} \cdots TE_{1m}$ modes all exist in degenerate pairs and all have a purely transverse magnetic intensity at the guide center-line; for these modes a ferrite rod on the axis may be used to produce Faraday rotation. The other TE_{nm}° modes of round guide* also exist in degenerate pairs, and Faraday rotation may be produced; however, in this case the proper location for the ferrite is not on the waveguide axis because there is no transverse magnetic intensity at that point. However, the transverse magnetic intensity at points off the guide axis is purely radial at certain angular positions,

POWER DIVISION

A — 100% → B

B < 50% → A
    50% → C

C — 100% → D

D < 50% → A
    50% → C
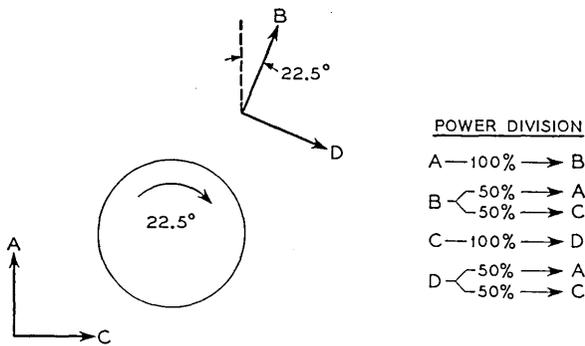
Fig. 20 — Non-reciprocal power divider using Faraday rotation.

* $n \neq 0$.

and purely circumferential at other angular positions. This is illustrated in Fig. 21 for the TE_{21} mode. The coupling which produces Faraday rotation is therefore from the radial magnetic intensity of one TE_{nm} mode to the circumferential magnetic intensity of the orthogonal degenerate mode. This coupling will be a maximum at the fractional guide radius where the product of the radial and circumferential magnetic intensities is a maximum — namely, where the following expression is a maximum:

$$\frac{R}{\rho} J_n{}' \left(\frac{k_{nm}{}'\rho}{R}\right) J_n \left(\frac{k_{nm}{}'\rho}{R}\right) \tag{17}$$

$$\rho = \text{radial coordinate}$$

$$R = \text{waveguide radius}$$

$$k_{mn}{}' = \text{the } m^{\text{th}} \text{ positive root of the Bessel}$$

$$\text{function derivative } J_n{}'(x) = 0$$

For the TE_{21} mode, the maximum coupling occurs at a fractional guide radius of 0.6. The optimum geometry of ferrite is a thin-walled tube of mean radius $0.6R$, longitudinally magnetized and centered on the waveguide axis.

5. TRANSVERSE FIELD EFFECTS IN RECTANGULAR WAVEGUIDE

We turn now to a consideration of reciprocal and non-reciprocal effects due to ferrite loading of a single-mode waveguide. Because there
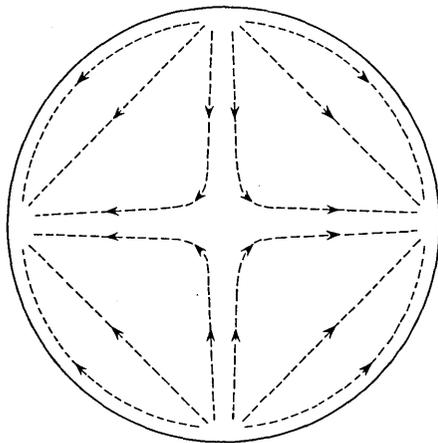
Fig. 21 — Pattern of magnetic intensity for the $TE_{21}0$ mode.

exists only one mode of propagation in each direction in the ferrite-loaded guide, the circuits using these effects are fundamentally different from those depending on Faraday rotation.

It is familiar to the art that every mode in a waveguide is characterized by three distinctive features;

(1) phase constant

(2) attenuation constant

(3) magnetic and electric field configurations in the waveguide

We will now show that suitable addition of magnetized ferrite may cause non-reciprocal as well as reciprocal changes in any of these distinguishing characteristics. It follows that the ferrite-loaded medium may be non-reciprocal by virtue of differing phase constants for the two directions of transmission, differing attenuation constants for the two directions of transmission, or differing field configuration for the two directions of transmission.

The configuration of magnetic intensity for the dominant wave in empty rectangular guide is sketched in Fig. 22(a). The H-loops lie
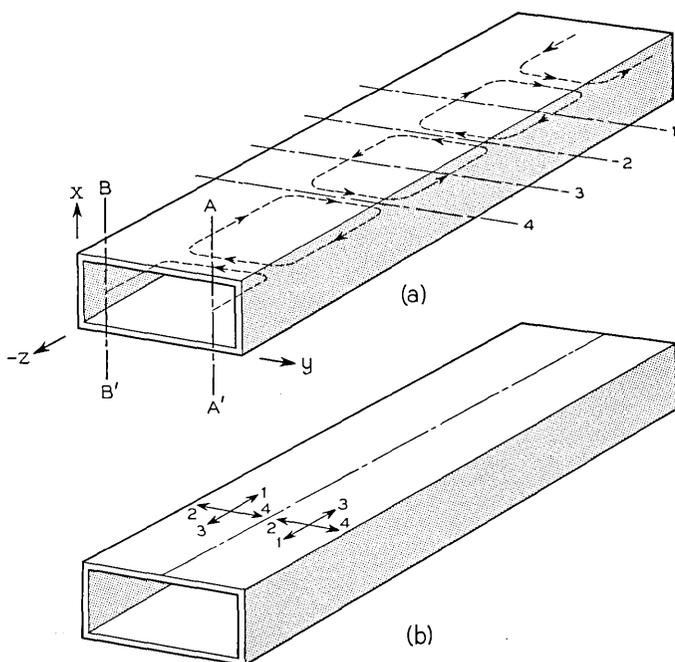
Fig. 22 — Magnetic intensity relations for the $TE_{10}\square$ mode showing the presence of elliptically polarized h.

entirely in planes which are parallel to the wide faces of the guide. At
any point other than at the center-line or at the side-walls of the guide,
the magnetic intensity vector is elliptically polarized in planes parallel
to the wide faces. This may be illustrated qualitatively by the series of
vectors intercepted by transverse planes 1, 2, 3, 4 of Fig. 22(a). A viewer
looking along the line A to A' will observe, for propagation in the $+z$
direction, the magnetic intensities at planes 1 through 4 in the time
sequence 1, 2, 3, 4; the spatial orientation of these vectors is sketched in
Fig. 22(b). It is evident that a viewer looking along A–A' will see a
clockwise rotating vector for wave propagation in the $+z$ direction.
For wave propagation in the $-z$ direction, the viewer will observe the
vectors in the time sequence 4, 3, 2, 1, and the viewer looking along
A–A' will see a counterclockwise rotating vector.

A viewer looking from B to B' will see a similar (but different) series
of vectors. The spatial orientations of the H vectors at the planes 1
through 4 are shown in Fig. 22(b) on the left-hand side of the guide
center-line. For wave propagation in the $+z$ direction the vectors appear
at B–B' in the time sequence 1, 2, 3, 4 and a counter-clockwise rotating
vector is observed. For wave propagation in the $-z$ direction, the
vectors appear at B–B' in the time sequence 4, 3, 2, 1, and a clockwise
rotating vector is observed.

Let us consider now the effect of introducing a sheet of ferrite, as
sketched in Fig. 23, sufficiently thin so that the fields are not appreciably
altered. The dielectric constant of the ferrite will perturb equally the
phase constants for the two directions of propagation, but the per-
meability of the ferrite will perturb unequally the phase constant for the
two directions. With the direction of magnetization shown in Fig. 23,
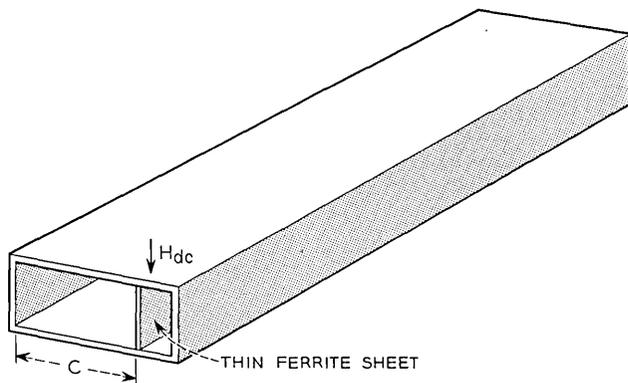
Fig. 23 — Non-reciprocal ferrite-loaded rectangular guide.

the electron spins are lined up perpendicular to the rf magnetic intensity, and thus the change in permeability introduced by the magnetized ferrite may be predicted using the known permeability values for the two senses of circular polarization (Fig. 6). With the ferrite on the right-hand side of the guide as shown in Fig. 23, and for wave propagation in the $+z$ direction, the positive circularly polarized component of rf h is greater than the negative circularly polarized component. Fig. 6 shows that the low-field permeability of the ferrite is less than unity under this condition. For wave propagation in the reverse direction the negative circularly polarized component predominates at the ferrite, and the low-field permeability of the ferrite is greater than unity. Thus, an increase of H_{dc} from zero (for which the ferrite-loaded guide is reciprocal) to values in the vicinity of saturation, introduces a decrease in phase constant for propagation in the $+z$ direction and an increase in the phase constant for propagation in the $-z$ direction. This configuration constitutes a *directional phase shifter*, whose total phase difference is proportional to the length of the ferrite-loaded medium, and when this difference is 180° the device is a *gyrator*.[19]

Suppose we introduce two identical thin ferrite sheets at equal distances from and parallel to the side walls as in Fig. 24. The above description of the fields illustrates that the static biasing field H_{dc} must be in opposite directions in order that the predominating rf h have the same sense of circular polarization at both ferrite elements. When biased oppositely as in Fig. 24, the two ferrite pieces cooperate in producing non-reciprocal phase shift.

If the static biasing field has the same magnitude and direction at
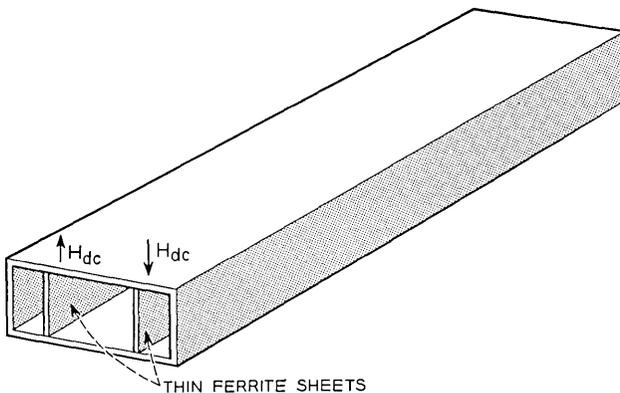
Fig. 24 — Alternative non-reciprocal ferrite loading for rectangular guide.

both ferrite elements of Fig. 24, the non-reciprocal phase contributions of the two elements cancel each other. However, a reciprocal phase change due to the magnetic field remains. This effect is small in the region where μ_+ and μ_- have approximately odd symmetry about the zero field point of Fig. 6. To the extent that the symmetry is perfect, the increase in μ_- should cancel the decrease in μ_+. For large applied fields or for operating frequencies near ferromagnetic resonance, the odd symmetry no longer holds, and the difference between μ_+ and μ_- will leave a reciprocal phase contribution. Calculations which clarify these interrelations are reported in Section 6.

The configuration of Fig. 24 is potentially non-reciprocal in another sense. For both elements biased downward and for wave propagation out of the paper, the right-hand ferrite element presents a permeability greater than unity and the left-hand ferrite element presents a permeability less than unity. As a consequence of this higher index of refraction on the right-hand side of the guide, one would expect a concentration of energy on the right-hand side of the centerline for wave propagation out of the paper. On reversing the direction of wave propagation, the sense of the predominating circularly polarized rf h is reversed at each ferrite element, the index of refraction is larger on the left-hand side of the guide, and one would expect to find the energy concentrated on the left-hand side of the guide. This situation is similar to the explanation for the asymmetry in the low-field loss of longitudinally-biased ferrite rods in circular guide, Section 4-8a. Experiments have shown that this *spatial displacement of the fields* does indeed take place, and we will presently describe how the effect may be incorporated into non-reciprocal devices. It should be noted that this non-reciprocal *field-displacement effect* may be obtained in configurations which have reciprocal attenuation and phase characteristics (such as Fig. 24), or in configurations which have non-reciprocal attenuation and phase characteristics (such as Fig. 23). In the circuit of Fig. 23, the redistribution of fields in the presence of the magnetized ferrite (the field-displacement effect) may be thought of as augmenting the differential phase shift which has been described above soley in terms of the difference between μ_+ and μ_- in an unperturbed $TE_{10}^{\square}$ field distribution.

Non-reciprocal attenuation in the ferrite-loaded medium may be realized using the configuration of Fig. 23. The biasing field H_{dc} is set at the value required for ferromagnetic resonance, H_{res} of Fig. 6, and since this resonance loss occurs only for the positive circularly polarized component, the transmission loss of the medium is larger for propagation in the $+z$ direction than for propagation in the $-z$ direction. The ratio

of these two losses is quite high, thereby forming a *resonance isolator*,*
when the ferrite is placed at a point where the ratio of the positive to
the negative circularly polarized components of h is high. The optimum
ferrite location for the resonance isolator and several other quantitative
aspects of the ferrite-loaded rectangular waveguide may be made clearer
with reference to the mathematical expressions for the fields.

The two linear components of magnetic intensity h_y and h_z for TE_{10}
in hollow rectangular guide, Fig. 22(a), are represented by the following
relations for propagation in the $+z$ direction:

$$h_y = D \sin\left(\frac{\pi y}{a}\right) \epsilon^{j(\omega t - (2\pi z/\lambda_g))} \tag{18}$$

$$h_z = \frac{D}{\sqrt{\left(\frac{2a}{\lambda_0}\right)^2 - 1}} \cos\left(\frac{\pi y}{a}\right) \epsilon^{j(\omega t - (2\pi z/\lambda_g) - \pi/2)} \tag{19}$$

where

$$D = \sqrt{\frac{2\pi}{\omega\mu\lambda_g ab}} \tag{20}$$

in which a, b = the large and small dimensions of the waveguide re-
           spectively

    λ_0 = free-space wavelength

    λ_g = guide wavelength

These expressions have been normalized for unit power flow in the di-
rection of propagation and the conductivity of the walls is assumed to
be infinite. Since the h_z and h_y components are in space and time quad-
rature, pure circularly polarized h exists when their magnitudes are
equal, a condition governed by the relation:

$$\left| \tan\left(\frac{\pi y}{a}\right) \right| = \frac{1}{\sqrt{\left(\frac{2a}{\lambda_0}\right)^2 - 1}} \tag{21}$$

A plot of the magnitudes of the h_y and h_z components as a function
of y/a is given in Fig. 25. It is seen that the location of the planes of
pure circularly polarized magnetic intensity depends upon the ratio of
the operating frequency f to the cut-off frequency f_c, approaches the
center line for a waveguide near cut-off, and approaches the side wall for
a waveguide remote from cut-off.

---

* The first published description of this form of isolator is in Reference 19.
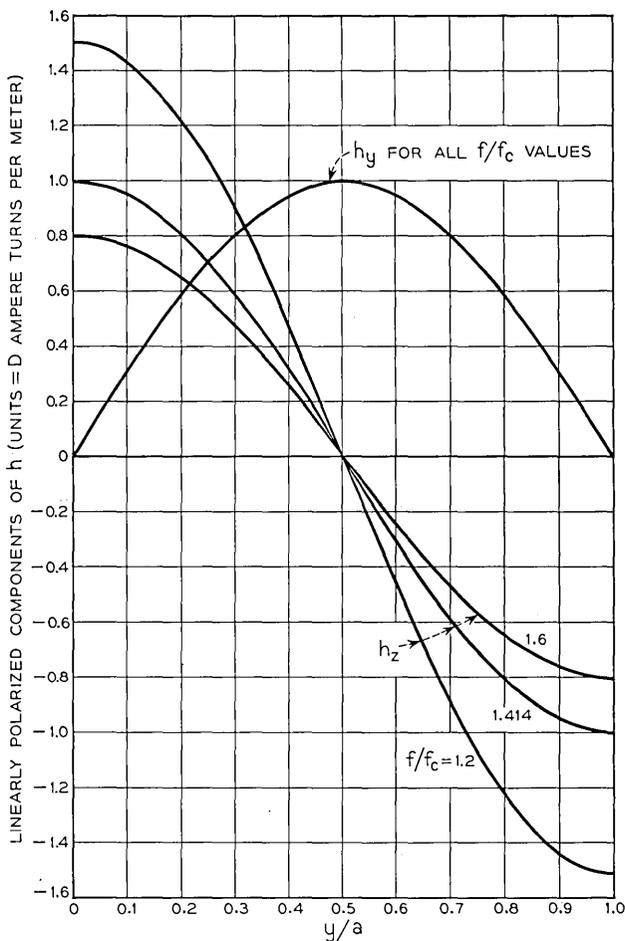
Fig. 25 — Theoretical magnitudes of transverse and longitudinal magnetic intensity of TE_{10} in empty rectangular guide.

For propagation in the $-z$ direction, h_y and h_z of (18) and (19) are altered by replacing

$$e^{-j(2\pi z/\lambda_g)} \qquad \text{by} \qquad e^{+j(2\pi z/\lambda_g)}$$

and by reversing the algebraic sign of h_z.

Since the ferrite presents a scalar permeability to circularly polarized waves, it is preferable in some instances to deal with the circularly polarized components of h in the hollow guide. These circularly polarized

components of h in the yz-plane are, for propagation in the $+z$ direction with reference to a magnetic field in the $-x$ direction:

$$h_{yz}{}^{\pm} = \frac{D}{2}\left[\sin\frac{\pi y}{a} \mp \frac{1}{\sqrt{\left(\frac{2a}{\lambda_0}\right)^2 - 1}}\cos\frac{\pi y}{a}\right]e^{j(\omega t - (2\pi z/\lambda_g))} \qquad (22)$$

For propagation in the $-z$ direction $e^{-j(2\pi z/\lambda_g)}$ is replaced by $e^{+j(2\pi z/\lambda_g)}$ and the sign in front of $\cos(\pi y/a)$ becomes $\pm$ instead of $\mp$. Thus, reversal of the direction of propagation reverses the sense of the circularly polarized component at a given point. Figure 26 shows the $h_{yz}{}^{\pm}$ components for $+z$ propagation and for several ratios of operating frequency to cut-off frequency, and illustrates that the location of the plane of pure circularly polarized h, as desired in the resonance isolator, depends on operating frequency. It is to be emphasized that the above field relations for the empty guide are valid only in case of small perturbations. In practical ferrite devices, these relations may be profoundly altered by the high dielectric constant of the ferrite.

5-1. *Rectangular Waveguide Pertubations Due to the Presence of Dielectric Material*

In the preceding section it was assumed that the field configuration was not appreciably altered by the addition of the ferrite. For very small pieces of ferrite this approximation is adequate and leads to a correct conclusion as to the general nature of the effects resulting from the addition of biased ferrite. However, when one tries to optimize the performance of a device which utilizes the interaction between the ferrite and the electromagnetic field, it is advantageous to use as much ferrite as is tolerable in order to obtain a large ferrite effect. When too much ferrite is added, the relatively high dielectric constant of the ferrite makes the waveguide medium capable of propagating secondary modes which can seriously interfere with the desired transmission properties. Calculations have been made to determine the conditions for cut-off of the TE_{20} mode in a rectangular waveguide for several locations of the ferrite in the guide. It was assumed that the relative permeability of the ferrite is unity (the field-free condition). The equations used in these calculations can be readily obtained by a technique due to N. H. Frank.[20]

Fig. 27 shows the results for the case of a longitudinal slab of dielectric of thickness d in the center of the guide as well as for dielectric of thickness $d/2$ at both sides of the waveguide. These two distributions of ferrite produce identical TE_{20} cut-offs because the dielectric is inserted
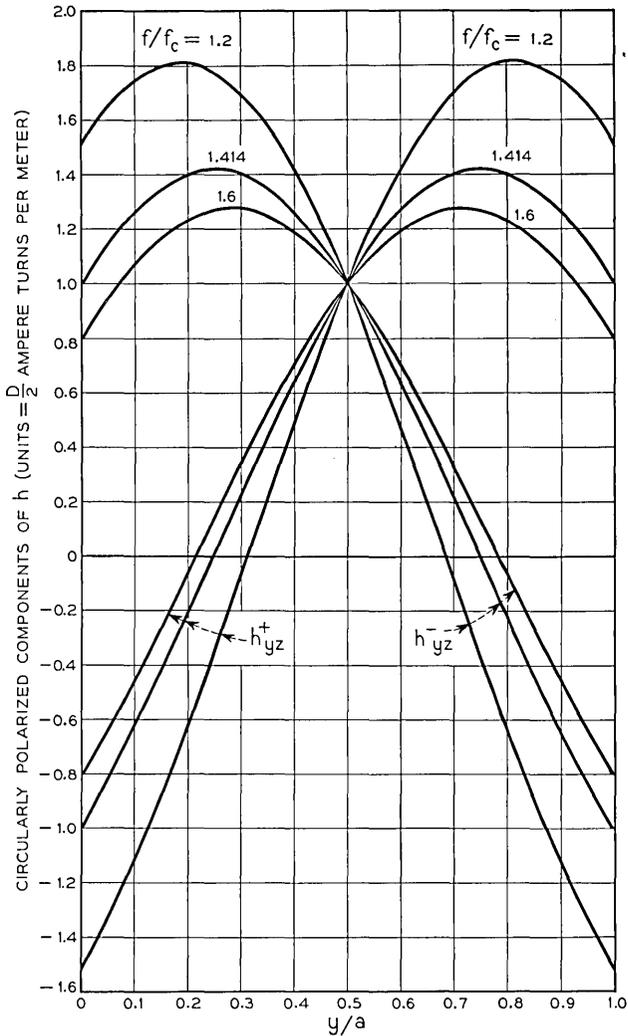
Fig. 26 — Theoretical magnitudes of circularly polarized h in empty rectangular guide.

in corresponding regions of field in both cases. One might presume that the curves would have a maximum slope when the transverse electric field maxima are passing through the interfaces between the dielectric and air. It should be noted, however, that the maximum slope occurs at values of d/a considerably less than 0.5, indicating that the field is
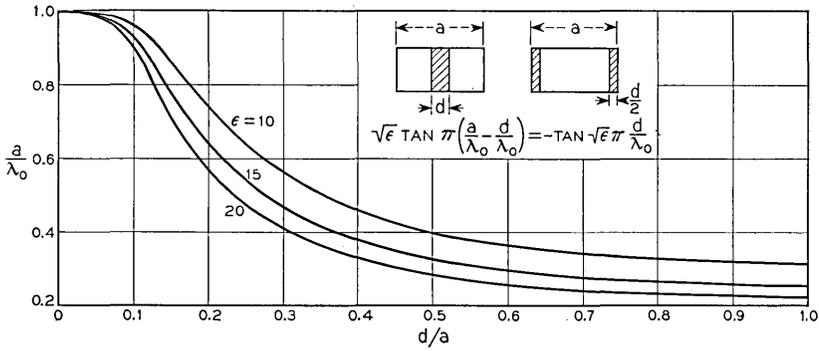
Fig. 27 — TE_{20} cutoff for rectangular guide with dielectric loading at the center or at both side walls.

being pulled into the dielectric. This field concentration in the ferrite is greater for increasing values of ferrite dielectric constant. The cut-off equation for these cases is:

$$\sqrt{\epsilon} \tan \pi \left(\frac{a}{\lambda_0} - \frac{d}{\lambda_0} \right) = -\tan \sqrt{\epsilon} \pi \frac{d}{\lambda_0} \tag{23}$$

Fig. 28 shows the curves of a/λ versus d/a for the case of a dielectric slab at one side of the guide. The equation for cut-off calculations is

$$\sqrt{\epsilon} \tan 2\pi \left(\frac{a}{\lambda_0} - \frac{d}{\lambda_0} \right) = -\tan \sqrt{\epsilon} 2\pi \frac{d}{\lambda_0} \tag{24}$$

Here one can observe that the curves go through two high slope regions
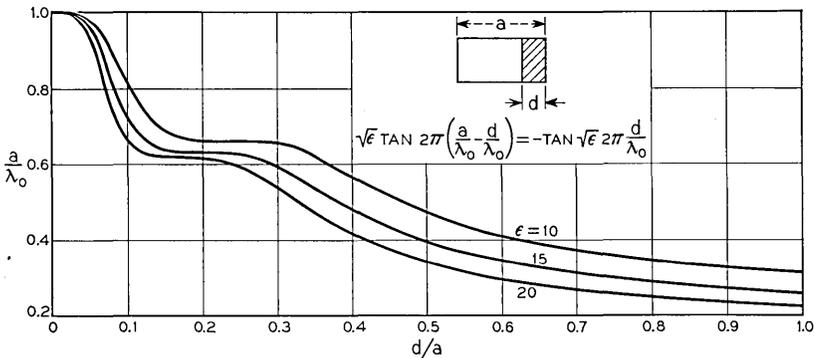
Fig. 28 — TE_{20} cutoff for rectangular guide with dielectric loading at one side wall.

since the transverse electric field of the TE_{20} mode has two maxima which pass through the air dielectric interface at two values of d/a. Again, because of the electric field concentration in the high dielectric constant material, these values of d/a are considerably less than 0.25 and 0.75, the values for maximum field in an empty guide. For small values of d/a, a smaller guide width will propagate the TE_{20} mode for this case of dielectric at one side than for the symmetrically placed dielectric as one might expect from the electric field distribution in the two cases.

Fig. 29 shows the results for a dielectric slab placed one quarter of the way from the side wall, where the dielectric should be most effective in propagating the TE_{20} mode, since it is in the maximum electric field region even at very small thickness. The cut-off equation is given by

$$\sqrt{\epsilon}\, \tan \pi \left(\frac{3a}{2\lambda_0} - \frac{d}{\lambda_0} \right) + \tan \sqrt{\epsilon} 2\pi \frac{d}{\lambda_0}$$

$$+ \left[\sqrt{\epsilon}\, \tan \pi \left(\frac{a}{2\lambda_0} - \frac{d}{\lambda_0} \right) \right] \qquad (25)$$

$$\left[1 - \sqrt{\epsilon}\, \tan \pi \left(\frac{3a}{2\lambda_0} - \frac{d}{\lambda_0} \right) \tan \sqrt{\epsilon} 2\pi \frac{d}{\lambda_0} \right] = 0$$

Equations for obtaining the phase constant of the TE_{10} mode in the dielectric-loaded waveguide, per Fig. 28, are available in the literature,[20] and the technique referred to above[20] can be employed to determine the TE_{10} phase constants for the configurations of Figs. 27 and 29.

The above theoretical relations are helpful in avoiding multi-mode conditions in ferrite-loaded rectangular waveguide. However, a very
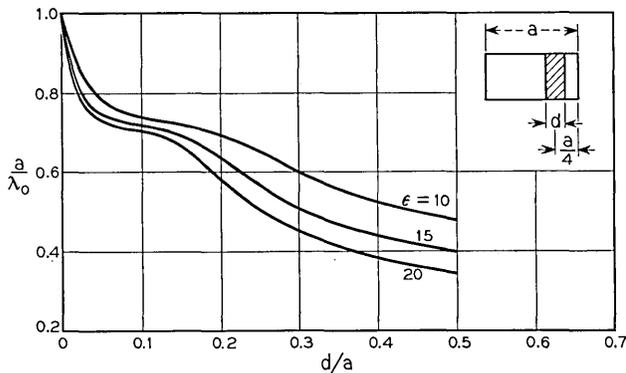
Fig. 29 — TE_{20} cutoff for rectangular guide with dielectric centered midway between the side wall and the center line.

effective means of increasing the available ferrite interaction before the onset of higher modes is to leave an air gap between the wide faces of the guide and the ferrite. Such an air gap introduces appreciable impedance to the displacement currents and allows the use of more ferrite. Unfortunately, the theoretical relations giving TE_{20} cutoff and TE_{10} phase constant under these conditions are not available.

5-2. *Rectangular Waveguide Directional Phase Shifters*

A simple gyrator in rectangular dominant-mode waveguide may be built using the configuration of Fig. 23 or 24, perhaps modified with an air gap between the ferrite sheet and the wide faces of the waveguide. The height of the ferrite sheets may be tapered to a point or stepped in a series of quarter-wave transformers in order to provide a smooth impedance transition from the empty waveguide to the ferrite-loaded region.

Theoretical work by H. Suhl and L. R. Walker[21] has shown that that optimum spacing between the side-wall and a very thin ferrite sheet to produce a maximum of differential phase shift for the two directions of propagation is $1/4$ of the guide width. A physical explanation for this optimum ferrite location can be given in terms of the point-field concept. The non-reciprocal phase shift due to the ferrite may be thought of as the result of a vector addition of the directly transmitted component of h with the reradiated component of h which is a consequence of the precessional motion of the ferrite's magnetization. From this point of view, the transverse h of the driving wave causes a reradiation via a longitudinal component of b and, as discussed in connection with equation (15), the effect will be proportional to the product of the transverse h and the longitudinal b. Since in this case both the driving and reradiated waves are dominant modes in the rectangular guide, the product of transverse h and longitudinal b is related to the ferrite location at point y by [see equations (18) and (19)].

$$\sin\left(\frac{\pi y}{a}\right)\cos\left(\frac{\pi y}{a}\right)$$

which is a maximum at $y = a/4$ or $3a/4$. Similar consideration of the longitudinal h of the driving wave results in the same conclusion.

For thicker sheets of ferrite, the optimum position for the ferrite is closer to the adjacent waveguide wall, due to a perturbation of the distribution of field components due to the dielectric constant of the ferrite. A plot of the measured difference in phase shift for the two directions of
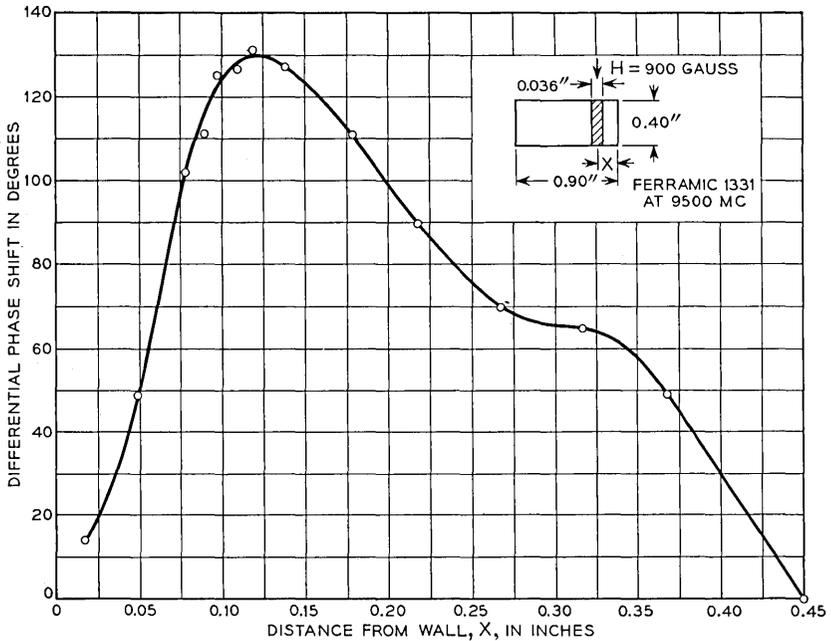
Fig. 30 — Measured non-reciprocal phase shift for a ferrite-loaded rectangular guide.

transmission as a function of the distance between the ferrite and the side wall is given in Fig. 30; even though the ferrite thickness is only 4 per cent of the guide width, the maximum differential phase shift occurs at a spacing of 0.133 times the guide width rather than at 0.25, as predicted by the perturbation theory.

A gyrator of the general form sketched in Fig. 23 has been built in the frequency region near 9,000 mc, and the observed difference in phase shift for the two directions of propagation as a function of frequency is shown in Fig. 31. If a smooth curve is drawn through the points shown, one would conclude that the differential phase shift is within ±10° of 180° over a 1,000-mc band. However, we believe the ripples in the curve as drawn are characteristic of the true performance, the accuracy of measurement being about ± 2°. These ripples cannot be due to multiple reflections of the dominant wave between the ends of the ferrite sheet, because the tapers are gradual enough to prevent reflections of the magnitude this would imply.

The TE_{20} mode is beyond cut-off in the ferrite region (on the presumption that the ferrite dielectric constant is less than 20), so we would not

expect mode conversion difficulties. Nevertheless, there is some indication that these ripples are due to higher order modes in the ferrite-loaded region, because a similar gyrator designed for a mid-band frequency near 11 kmc in the same (0.400″ x 0.900″) waveguide showed appreciably larger phase ripples.

An alternative form for the non-reciprocal phase shifter consists of a series of rods of ferrite extending entirely through the waveguide in the plane of the ferrite sheets of Figs. 23 and 24. In this form there need be no air gap between the magnetic structure used to bias the ferrite and the ferrite itself because the rods can extend through the waveguide wall without radio frequency power leakage.

5-3. *Rectangular Waveguide Circulators Using Gyrators*

With the development of the rectangular waveguide directional phase shifter, it becomes possible to build a variety of rectangular waveguide circulators by combining this new circuit element with standard directional couplers.[8,22]

One of the simplest such circulator configurations is shown in Fig. 32. This circulator consists of two 3 db (0.707 amplitude) couplers joined by sections of guide so that one guide has π radians more phase shift for transmission from left to right in the direction of the arrow than for the opposite direction of propagation. This non-reciprocal phase shift is provided by a piece of ferrite of appropriate length near the side wall of the rectangular guide with an appropriate dc magnetic field. The other guide must have a dielectric counterpoise to balance the reciprocal phase shift provided by the ferrite in the bottom guide. One can easily show that the above configuration will give a circulator action.
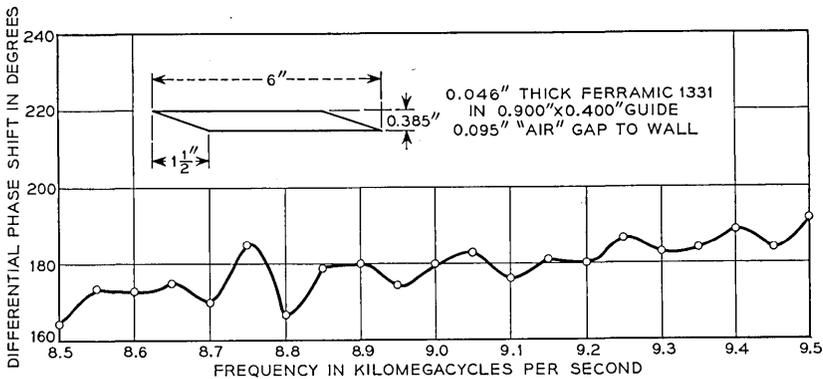
Fig. 31 — Frequency characteristic of a 9,000-mc gyrator.

This type of circulator has been built for operation at 9,000 mcs using two Riblet couplers with the magnetic field of about 1,000 oersteds provided by a permanent magnet. The overall length of the device was 12 inches. The forward loss was about 0.1 db while the signal in the "zero-transmission" terminals was down by at least 30 db. No bandwidth data have as yet been taken.

The above scheme looks very similar to the circulator using two hybrids and a gyrator which has been described by C. L. Hogan. However, this circulator design can be generalized in that a number of different configurations can be designed, some of which may have different and useful properties.

Thus, Fig. 33 shows an alternative circulator scheme which uses two 90° directional phase shifters, one in each line. An extra quarter wavelength reciprocal section is required in one of the guides as shown. This scheme should shorten the overall length of the circulator, but it has been found to be more difficult to adjust because the magnetic field used to balance the non-reciprocal phase shift also affects the reciprocal phase shift.

The circulator shown in Fig. 32 depends for its action on the fact that the ferrite gives a non-reciprocal phase shift of exactly 180°. Any departure from this value will cause some undesirable transmission in the "zero-transmission" direction. However, the amount of phase shift given by the ferrite is dependent on the temperature of the ferrite and the frequency of operation. Therefore, the directional-phase section may have a phase shift differing from π by a small angle, Δ, at a temperature or a frequency differing from the design value.
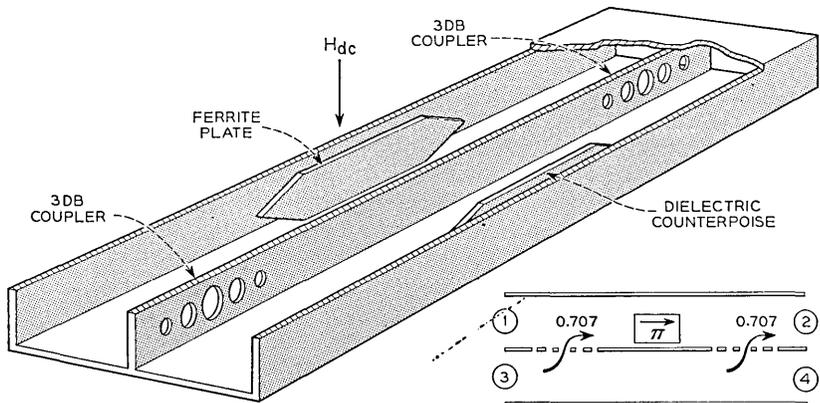
Fig. 32 — Circulator using 3-db couplers and a gyrator.

In order to minimize the effect of this change in the value of the phase shift, we shall show that one can split the coupling sections in a symmetrical and appropriate manner. Fig. 34 shows such a three-section circulator which minimizes the effects of temperature or frequency variations. The necessary values of coupling in this three-section circulator can be determined as follows: The amplitude coupling in the first and last couplers is equal to $\sin \theta$, while the two middle couplers each have a coupling equal to $\sin \varphi$. In order that complete coupling take place in the desired direction, $2\theta + 2\varphi$ must equal 90°.*

From the above it may be shown that $\cos 2\theta = \sin 2\varphi$ and $\sin 2\theta = \cos 2\varphi$. Let us assume that all the $\vec{\pi}$ sections have a phase shift of $\overrightarrow{\pi + \Delta}$. With an input of unit amplitude at (3) in Fig. 34, let us calculate the output at (2). The following table shows how one proceeds, taking into account only differences between the phase shifts occurring in top and bottom guides:

	At a	At b	At c	At d
Upper guide..	1	$\cos \theta$	$\cos \theta$	$\cos \theta \cos \varphi + \sin \theta \sin \varphi \, e^{j\Delta}$
Lower guide..	0	$j \sin \theta$	$-j \sin \theta \, e^{j\Delta}$	$-j \sin \theta \cos \varphi \, e^{j\Delta} + j \sin \varphi \cos \theta$

and so on until we reach the output at (2). This is given by
$$\sin \theta \cos \theta \cos^2 \varphi [1 - e^{j3\Delta}] - [\cos^2 \theta \sin \varphi \cos \varphi$$

$$- \sin^2 \theta \sin \varphi \cos \varphi - \cos \theta \sin \theta \sin^2 \varphi] [e^{j\Delta} - e^{j2\Delta}] \quad (26a)$$

which we want to depart from zero at a minimum rate. We thus have an expression which we shall write for simplicity as
$$A[1 - Be^{j\Delta} + Ce^{j2\Delta} - De^{j3\Delta}]$$

It is well known that if 1, B, C, D form a binomial relation, 1-3-3-1,
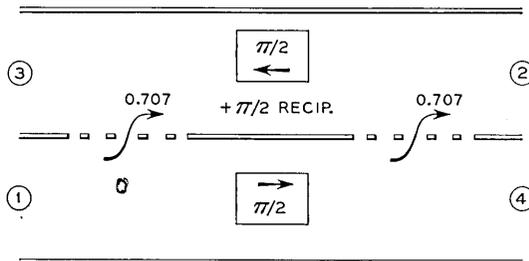
Fig. 33 — Alternative form for circulator of Fig. 32.

* For further details on couplers see Miller, Reference 8, p. 694.
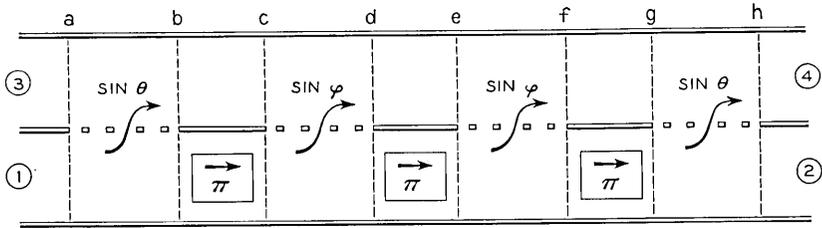
Fig. 34 — Circulator using three gyrators for broad banding.

then the above expression will be equal to $A(1 - e^{j\Delta})^3$, which is equal to $(j\Delta)^3 +$ higher order terms since*

$$1 - ne^{j\Delta} + \frac{n(n-1)}{1\cdot 2} e^{2j\Delta} + \cdots + (-1)^n \frac{n(n-1)\cdots 1}{1\cdot 2\cdot 3 \cdots n} e^{nj\Delta} \qquad (26b)$$

$$= (1 - e^{j\Delta})^n = (j\Delta)^n + \text{higher order terms}$$

Therefore, by equating corresponding terms in (26a) and (26b) we can solve for θ and φ. The result is $2\theta = 21.466°$ and $2\varphi = 68.534°$. With this design the unwanted output at (2) with input at (3) will therefore be proportional to $(i\Delta)^3$ which is, of course, much smaller than for the single section circulator for which the output at (2) is proportional to $j\Delta$. The above scheme can be used for designing an n section circulator which will be compensated for changes in the $\overrightarrow{\pi}$ phase shift to $(j\Delta)^n$.

With the rectangular waveguide gyrator, one can also make a variety of one-way couplers which couple only part of the power over in a non-reciprocal manner. Thus, Fig. 35 shows the rectangular waveguide equivalent of the non-reciprocal power divider described in Section 4-9c.
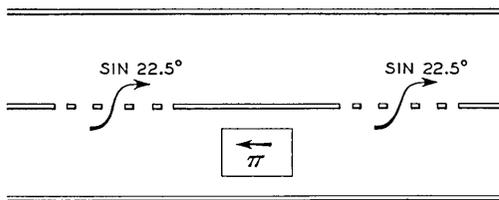
Fig. 35 — Non-reciprocal power divider in rectangular guide (analogue of Fig. 20).

* See Kyhl, Reference 8, p. 882.

5-4. *Circulators Employing Non-Reciprocal Phase Constants*

Two transmission lines having non-reciprocal phase constants may be employed to form a circulator as sketched in Fig. 36. The coupling apertures between guide A and guide B are predetermined in size and number so that complete transfer from one guide to the other will occur when the phase constants are equal.* If the magnetizing field in Fig. 36 were zero and the phase constants of the two guides were equal, then power entering at 1 would emerge entirely at 2, and of course reciprocity would hold. In the circulator application of this configuration the phase constants for the two guides at zero H_{dc} are adjusted to be different through a suitable choice of the waveguide width, thickness of ferrite, or position of ferrite in the two guides. As H_{dc} is increased from zero, the phase constants for the two guides vary in the manner sketched in Fig. 37. The forward waves in guides A and B have the same phase constant for the $H_{dc} = H_1$ whereas the backward waves in the two guides have different phase constants at this same value of H_{dc}. Under this condition, transmission will occur from terminal 1 to terminal 2 (Fig. 36) because the phase constants are equal and the number and size of coupling apertures has been preselected for complete power transfer. Power entering in terminal 2, however, will not return to terminal 1 because the phase constants are unequal. Components of power transferred from guide B to guide A at the various coupling holes add up destructively in guide A when the phase constants are unequal and for a suitably
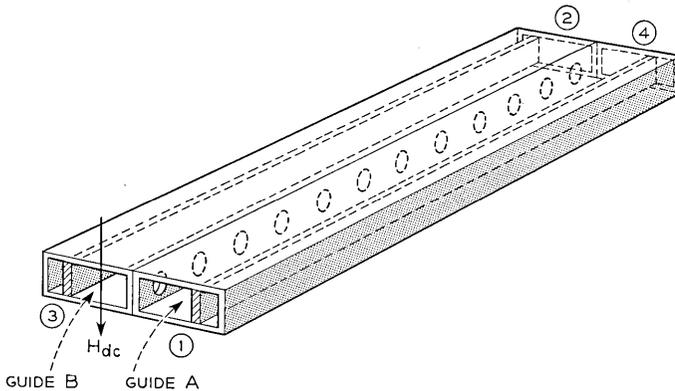
Fig. 36 — Rectangular waveguide circulator using distributed reciprocal coupling between waveguides having non-reciprocal phase constants.

* S. E. Miller, Reference 8.

large ratio of the difference between the phase constants to the coefficient of coupling between the guides, the power transferred from 2 to 1 can be made small. Power entering at terminal 2, therefore, is transferred to terminal 3. From the symmetry of the structure it is apparent that power entering at 3 will also be transferred to terminal 4 and power entering terminal 4 will return to terminal 1. Thus, we have the characteristics previously defined as those of a *circulator*.

The structure of Fig. 36 may be simplified by employing a section of ferrite in only one of the waveguides at the expense of some frequency selectivity due to dielectric asymmetry. In some applications this selectivity may be an advantage.

It may be noted that the equality of phase constants can be adjusted in a completed assembly by adjusting the strength of the magnetizing field.

The coupled-wave circulator may also be adapted to serve as a mode selective device. Fig. 38 shows one such structure, a transducer from single mode rectangular guide to multimode or single mode round waveguide. One may couple either to the dominant (TE_{11}), to the circular electric (TE_{01}), or to any other TE mode of round guide using the configuration sketched in Fig. 38, and by coupling through the wide wall of the rectangular waveguide the TM modes of round guide may be used. Let us assume that the phase constant of the rectangular guide is made equal in one direction to that of the TE_{01} mode in round guide
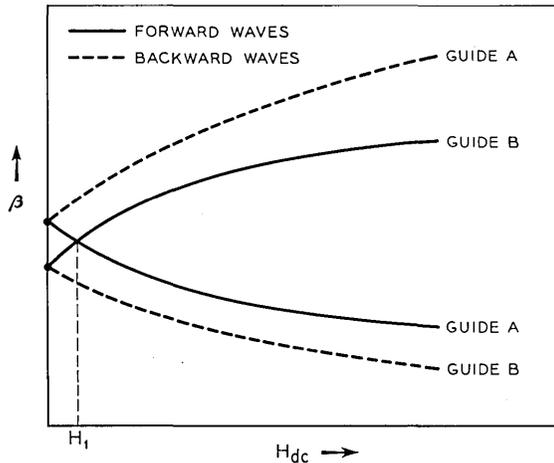
Fig. 37 — Phase constants vs applied field for the guides of Fig. 36.

and that the phase constant for the rectangular guide in the reverse direction is unequal to the phase constant of any one of the modes of the round guide. Then the transmission effects will be as follows:

(1) From terminal 4 to terminal 1 only
(2) From terminal 1 to terminal 2 in the TE_{01} mode only
(3) From terminal 2 in any mode to terminal 3 in the same mode
(4) From terminal 3 in any mode except TE_{01} to terminal 2 in the same mode. From terminal 3 in TE_{01} to terminal 4 only.

The significant feature of the configurations described in this section is the use of reciprocal coupling between transmission lines having non-reciprocal phase constants, and this idea may be generalized to other forms of non-reciprocal transmission line.

5-5. *Resonance Isolators*

We turn now to a device making use of the non-reciprocal attenuation of a ferrite-loaded guide. In Section 5 it was shown that a thin strip of ferrite placed in a rectangular guide, as in Fig. 23, forms a resonance isolator if the applied magnetic field is of the right magnitude to produce ferromagnetic resonance.

We have shown experimentally that for a properly positioned peice of ferrite large differences in transmission loss for the two directions of propagation can indeed be observed. Information on one such isolator designed for 4 kmc is given in Fig. 39. While the dc magnetic field required to produce resonance was of the order of ω/γ, it deviated appreci-
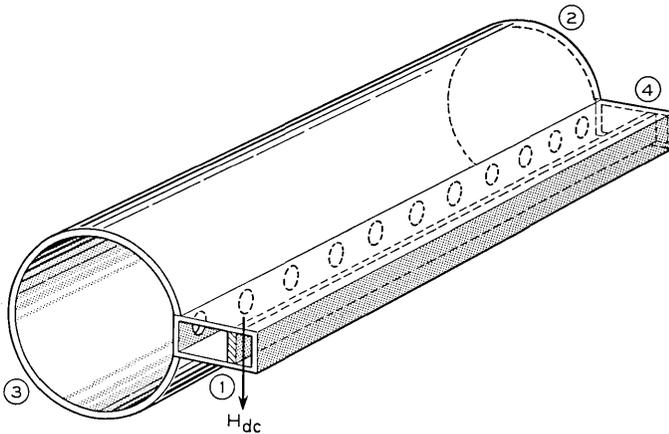
Fig. 38 — Mode-changing circulator analogous to Fig. 36.

ably from this value by virtue of the dc demagnetizing factor of the ferrite strip. Additionally, it was observed that the field required to produce maximum loss was also a function of the distance of the ferrite from the side wall of the waveguide. In other words, the circuit environment of the ferrite plays an important part in determining the resonant frequency.

Since there is no theory available for a ferrite-loaded waveguide operating near or at gyromagnetic resonance, the design of such isolators has been empirical. Many different geometries have been tried, but in general we find that the performance of this type of isolator is limited by an unexpectedly high forward loss. Thus, while enormously high reverse losses are easily obtainable using reasonably short ferrite-loaded sections, the presence of this forward loss has prevented us from observing reverse-loss to forward-loss ratios of greater than 20 to 1 or 25 to 1 in decibels, even for very thin ferrite slabs. This ratio of reverse loss to forward loss is approximately independent of length, and one can elect a forward loss of 1 db with a reverse loss of 20 db, or a forward loss of 4 db with a reverse loss of 80 db. Examination of the loss as a function of applied field for both directions of propagation yields curves of the type shown in Fig. 40. It is seen that for both the plus (forward) and minus (reverse) directions of propagation there is a resonant rise in the loss for particular values of field applied. In general, the loss peaks for the plus and minus directions do not occur at exactly the same value of applied field. The presence of the resonant loss for the minus wave is the reason for the limitation on back-to-front ratios.

This resonant loss for the forward direction is not easily explainable. As seen from Fig. 40, it is present for all positions of the ferrite slab. This forward loss would be expected for a thick ferrite slab, since not all of the ferrite can be at the point of pure circularly polarized h. How-
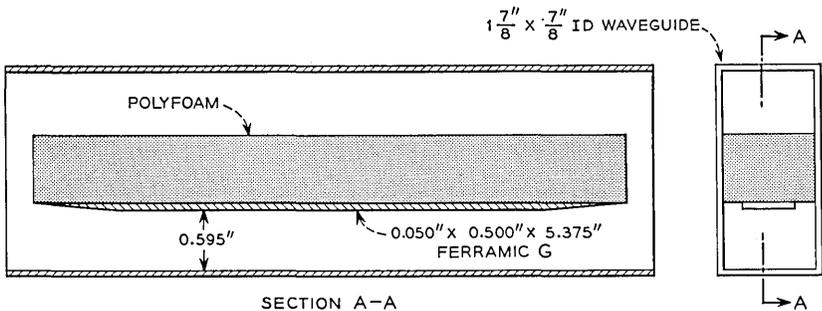
Fig. 39 — 4-kmc resonance isolator.
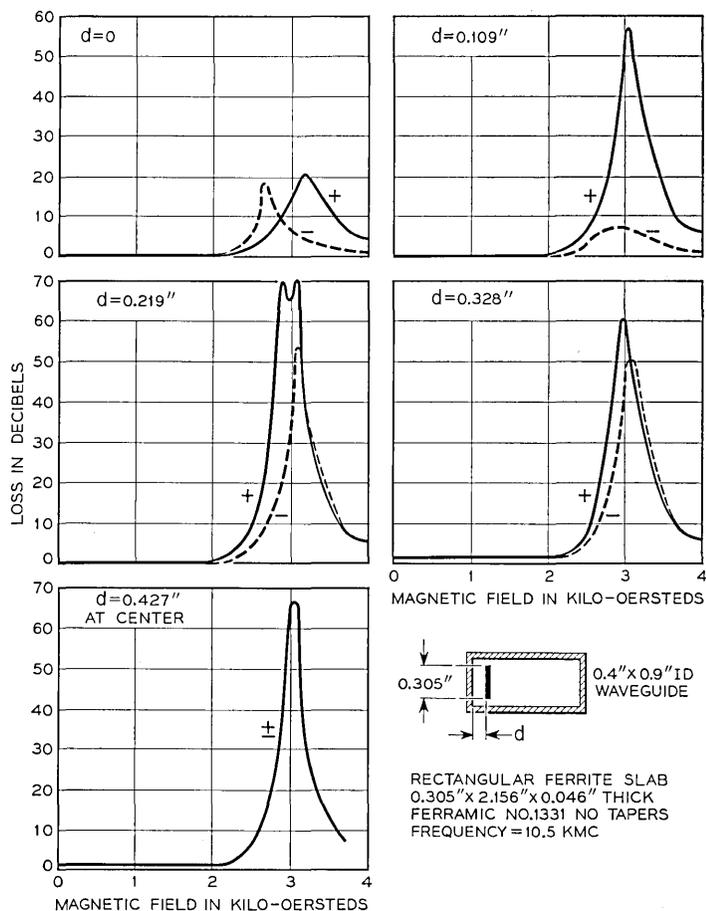
Fig. 40 — Resonance absorption for the forward (+) and reverse (−) directions of propagation as a function of the position of the ferrite element.

ever, if the ferrite is made thin enough and is placed at exactly the right point, one would not expect to find any plus polarized rf field in the ferrite when waves are passing in the minus direction through the waveguide, and the forward loss ought to remain low for all values of dc field. In spite of these considerations, however, it has not been our experience that making the ferrite slab thinner has eliminated the presence of this resonant loss for the minus direction of propagation. If the ferrite slab extends completely from top to bottom of the waveguide, the back-to-front ratio for transmission loss has been observed to be quite

poor and of the order of 8 or 10 to 1 in decibels. However, by cutting down the height of the ferrite slab to perhaps $\frac{2}{3}$ of the height of the narrow wall of the waveguide, the ratio of 20 or 25 to 1 has been obtained. This strongly suggests that dielectric effects may be playing an important part, since reduction in the height should reduce the displacement current through the ferrite much more rapidly than the interaction of the ferrite with the magnetic field.

In adjusting the position of the ferrite in the waveguide, it has been observed that the position which produces the highest reverse losses is not in general the same as the position which will produce the smallest resonant loss in the forward direction. This also may be seen from Fig. 40. In selecting the best compromise to give a high reverse-to-forward loss ratio, it has been found that the ferrite should always be placed so that the resonant loss in the forward direction is a minimum. In Fig. 40 this corresponds to the $d = 0.109''$ condition. Having so positioned the ferrite, it is also found that the magnetic field which produces the largest reverse loss is not the same as the field which produces the least forward loss. We observe that the best reverse-to-forward loss ratio is obtained when the magnetic field is adjusted for the maximum of reverse loss.

It is interesting to note that as shown by Fig. 40 the fields required to produce resonance in the $(+)$ and $(-)$ directions change as a function of displacement of the ferrite from the sidewall of the waveguide. The resonant frequencies are different for the two directions of propagation except when the ferrite is on the center line, when it becomes perfectly reciprocal.

Fig. 40 also demonstrates the typical behavior of a ferrite with low magnetic loss at zero field·strength. With the ferrite slab against the side wall, it is in a position of very low electric field, and any loss observed should be principally due to the rf magnetic field. Fig. 40 shows no observable loss at the sidewall position, and it is believed that for this ferrite the magnetic loss is less than the dielectric loss. Increase in zero field loss as the slab is moved toward the center is believed due to dielectric loss. However, many other ferrites show a substantial loss at the sidewall which decreases as the slab is moved toward the center line. These are believed to have zero field magnetic losses considerably higher than the dielectric losses.

Although a resonance isolator depends upon the adjustment of the magnetic field to produce a resonance condition, this device is not narrow-band if ferrites with broad ferromagnetic resonance lines are used. In Fig. 41 is shown the way in which the forward and reverse losses
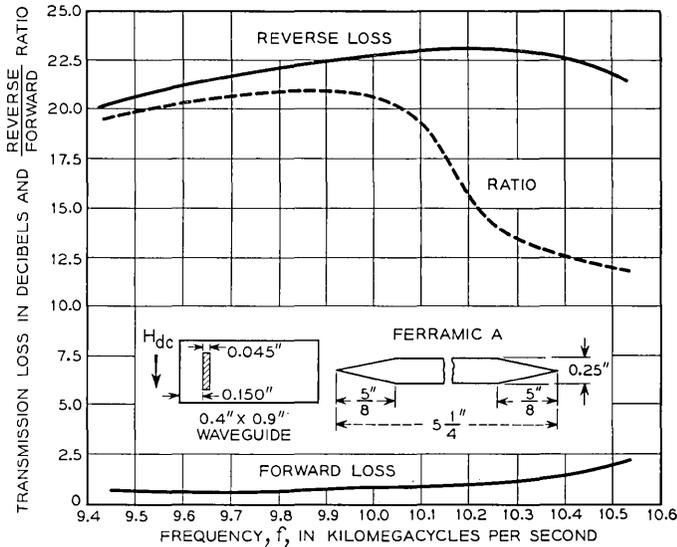
Fig. 41 — Frequency dependence for a 10-kmc resonance isolator.

vary in the frequency range from 9.5 to 10.5 kmc for the isolator de-
scribed in the same figure. This isolator was a laboratory model which
was lined up at 10 kmc, and no attempt was made to optimize the design
for a broad band. The curves shown are what resulted from the adjust-
ment procedure of the type described above. As may be seen, a reverse-
to-forward loss ratio of better than 20 to 1 in db is maintained over at
least a 5 per cent band.

The field of usefulness of the resonance isolator is probably limited
to frequencies below about 20 kmc. Above this frequency the magnetic
field required to produce resonance becomes large and inconvenient.
Below this frequency, however, fields of resonant strength are easily
obtained from Alnico magnets of convenient size, and the extreme
simplicity and compactness of the device makes it attractive. Other iso-
lators (to be described) have similar attractive characteristics.

5-6. *Field-Displacement Isolators and Circulators*

We turn now to devices which utilize the non-reciprocal field dis-
placement effects which occur in a ferrite-loaded guide:

(a) *Resistance-Sheet Isolator*

Fig. 42 illustrates a field displacement isolator which has already been
found attractive in practice. In this device the predominance of one

type of circularly polarized magnetic intensity at each of the ferrite sections results in a shift of the energy density and the associated electric field to one side of the center line for propagation in one direction and to the other side of the center line for propagation in the opposite direction, as qualitatively sketched in Fig. 42. These two distributions of electric field would result in appreciably different attenuations for forward and reverse directions of propagation, due to the addition of the resistance sheet shown adjacent to the left hand ferrite element in Fig. 42; such a difference in attenuation would of course result in isolator behavior.

Fig. 43 shows the observed performance of a resistance-sheet isolator (see Fig. 44) made for operation in the 24,000 mc region using 0.170″ x 0.420″ I.D. waveguide and 0.043″ x 0.162″ x 1.5″ sections of Ferramic J, quoted by General Ceramics Company as having a saturation magnetization of 2,900 gauss. Let us first note that ratios of reverse loss to forward loss as high as 30 were observed in this model. Several additional details of the observed performance merit recording in order to provide guidance for theoretical analysis. As would be expected, optimum performance was obtained when a maximum amount of ferrite was employed, the upper limit being determined by the onset of mode conversion. However, it was also observed that the resistance sheet isolator could employ more ferrite than could the same rectangular guide in the absence of the resistance sheet. For example, the 0.043″ thickness of ferrite employed in the model of Fig. 44 was adequate to cause mode conversion in the absence of the resistance sheet, such mode conversion
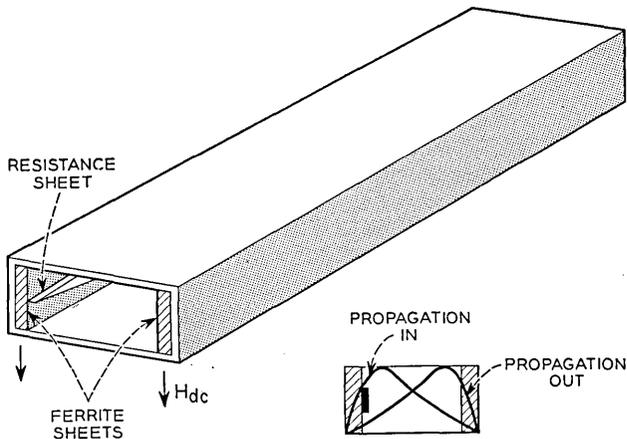
Fig. 42 — Resistance-sheet field-displacement isolator.
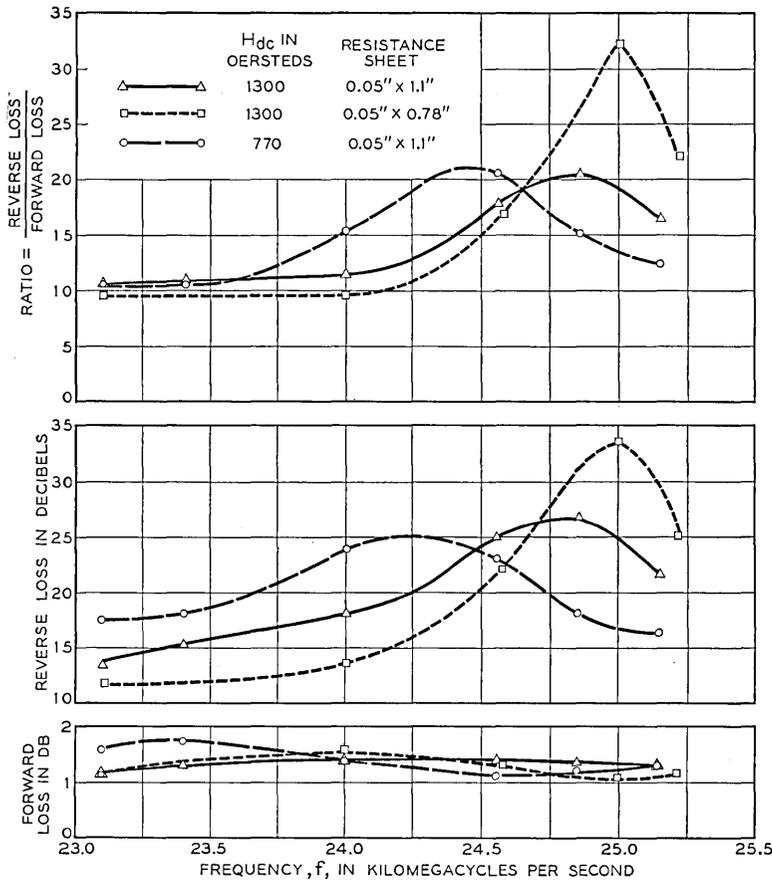
Fig. 43 — Observed performance of a 24,000-mc resistance-sheet isolator.

being characterized by fluctuation between zero loss and loss in excess of 5 db for varying magnetic field of either polarity. Evidently the presence of the resistance sheet damps out the higher-order modes. An increase in ferrite thickness to 0.053″ was adequate to cause mode conversion in the presence of the resistance sheet.

It is noteworthy also in connection with the data of Fig. 43 that the forward loss is approximately independent of frequency whereas the reverse loss is rather frequency sensitive.

The magnetizing field in the model of Fig. 44 was adequate to bring the ferrite near saturation, and in this region variations in field strength produced only small changes in observed performance. For example,

curves I and III apply to the identical isolator structure at applied fields of 1,300 and 770 oersteds, respectively.

One of the subtleties in the observed performance of the resistance sheet isolators is illustrated by comparison of curves I and II of Fig. 43. These two sets of data apply to the same configuration with the exception of the length of the resistance sheet. Both resistance sheets gave essentially the same forward loss, but the shorter sheet gave a more pronounced peak in the reverse loss, resulting in a higher reverse-loss to forward-loss ratio.

Another model of the resistance-sheet isolator was made in the 6-kmc region employing the parts shown in Fig. 45 comprising a 0.622″ x 1.372″ I.D. waveguide, a 0.003″ x 0.2″ x 2.48″ resistance sheet, a 0.225″ x 0.570″ x 2.97″ and a 0.196″ x 0.570″ x 2.12″ sections of magnesium aluminum ferrite having a saturation magnetization of about 1,800 gauss. (The latter ferrite pieces have odd dimensions because they were fabricated from pieces remaining from other experiments). The observed performance of this model is recorded in Fig. 46. For the data shown, an applied field of about 470 oersteds was employed, but a reduction of the applied field to 270 oersteds had only a minor effect on the observed data — namely, the peak ratio of reverse-loss to forward-loss remained
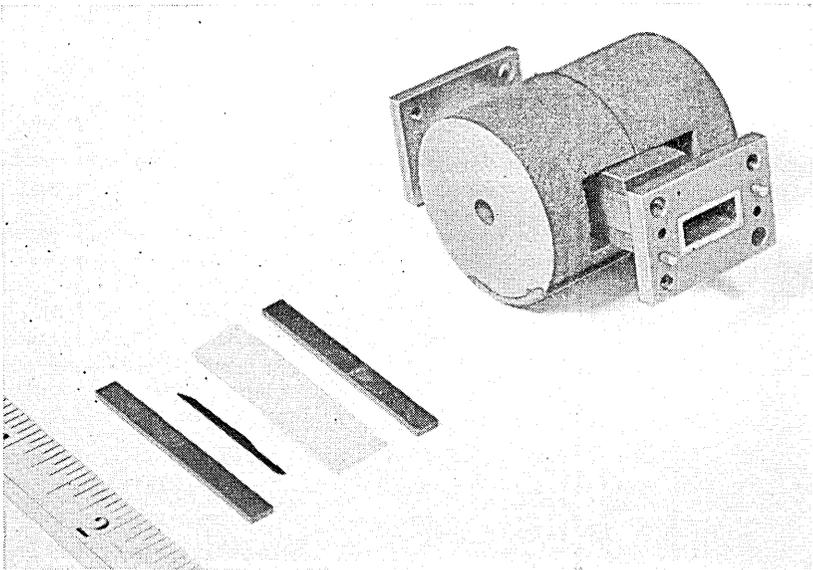
Fig. 44 — The resistance-sheet isolator used to obtain the data of Fig. 43.
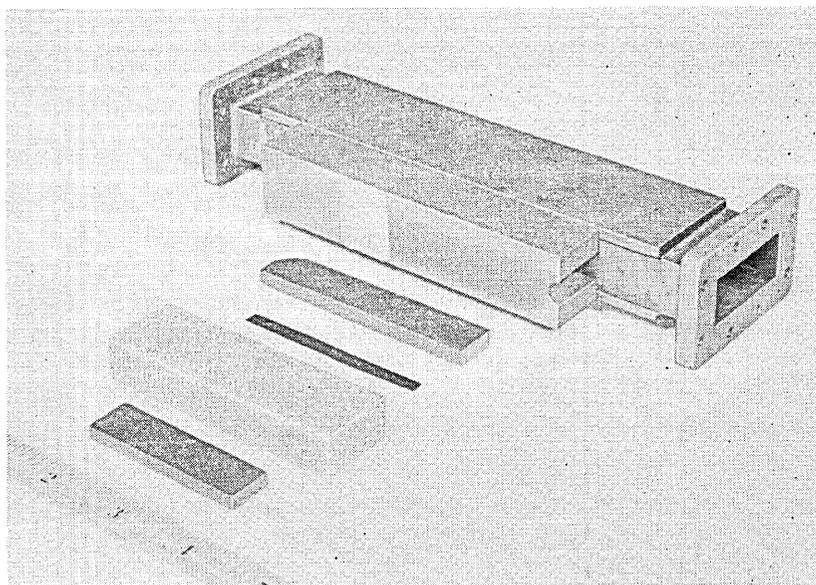
Fig. 45 — The resistance-sheet isolator used to obtain the data of Fig. 46.

80 to 1, but moved down to 6.1 kmc. Likewise the reduction in applied field translated the observed reverse-loss and forward-loss curves toward slightly lower frequencies.*

The return loss for the input and output terminals of the above isolators was in the region 26 to 29 db when the ferrite ends were untapered. The 24,000-mc model showed return losses in the range 33 to 37 db over the 23- to 25-kmc region using $\frac{3}{8}''$ long linear tapers at the ferrite ends.

The 6-kmc model was produced with small effort compared with that devoted to the 24,000-mc model of Figs. 43 and 44. It is likely, therefore, that there is an inherent reason why the resistance-sheet isolator behaves better in the lower frequency region.

When looking at the 0.3 db observed forward loss of the 6 kmc model in contrast to the reverse loss of greater than 20 db, one sees immediately that the electric field at the resistance sheet must be very small in the forward direction as compared to the electric field at the resistance sheet in the reverse direction. (There is no assurance, however, that the elec-

---

* Subsequent work by S. Weisbaum of Bell Telephone Laboratories has produced a resistance-sheet isolator having a reverse-loss to forward-loss ratio greater than 150 over a 500-mc band near 6,000 mc. A single ferrite element, spaced from the side wall, was used in these models.
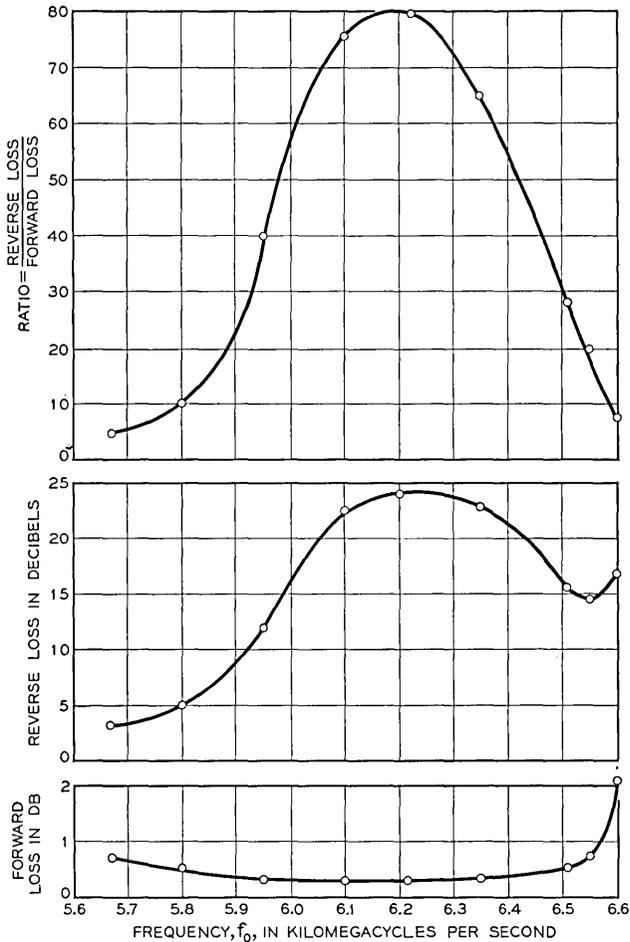
Fig. 46 — Observed performance of a 6 kmc resistance-sheet isolator.

tric field causing the reverse loss is a transverse component). Work done
by E. H. Turner of Bell Telephone Laboratories subsequent to that
described above has led him to conclude that a complete electric-field
null at the edge of the ferrite element may occur provided that the
ferrite's tensor permeability component k^2 is equal to or greater than
μ^2. The two isolators described above are operated at magnetizing fields
such that k^2 is small compared to μ^2 and Turner's condition for the null
is not met. However, the ferrite's k/μ for the 6-kmc model was larger
than that for the 24-kmc model, and since the 6-kmc model performed

better with less effort it is likely that the larger k/μ is desirable in field-displacement devices. The requirement of appreciable k/μ suggests the need for ferrites of higher saturation magnetization for field-displacement devices to be used at 24 kmc or higher frequencies.

(b) Slotted-Wall Field-Displacement Isolator

Another form of isolator depending on non-reciprocal field displacement is sketched in Fig. 47. This circuit depends upon the fact that a longitudinal slot may be cut in the wide face of a dominant-mode rectangular waveguide along a line of zero longitudinal magnetic intensity without appreciably altering its transmission characteristics. This is a familiar property which is utilized in standing wave detectors. The addition of magnetized ferrite in the manner sketched in Fig. 47 has the effect of displacing the null for the longitudinal intensity to one side of the center line for one direction of transmission and to the opposite side of the center line for the reverse direction of transmission. The slot of Fig. 47 is located so as to coincide with the null for the longitudinal magnetic intensity for one direction of propagation, and as a consequence there will be negligible attenuation for that direction of propagation. However, for propagation in the reverse direction a very appreciable value of longitudinal magnetic intensity is present at the slot with the result that radiation up into the lossy dielectric occurs.

Exploratory experiments were carried out at 24,000 mc using the structure of Fig. 47 modified to have ferrite only on the side of the waveguide toward which the longitudinal slot was displaced. Using a ferrite section 0.050″ x 0.162″ x 1.5″ and a somewhat shorter slot, a forward
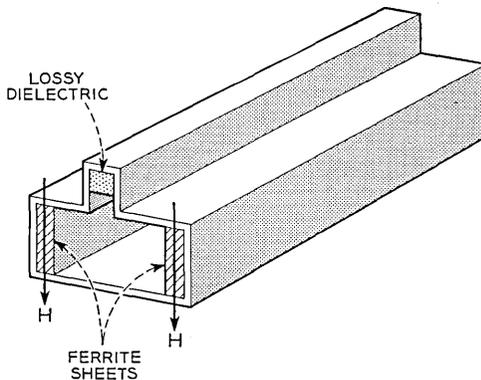
Fig. 47 — Slotted-wall field-displacement isolator.

loss of 0.6 db and a reverse loss of 20.5 db were observed at one frequency.
A frequency characteristic was not taken. It is noteworthy, however,
that the forward loss in the presence of the magnetizing field was rela-
tively independent of the strength of the magnetizing field, whereas the
reverse loss was critically dependent on the strength of the magnetizing
field. It is possible that the longitudinal dimension of the slot was near
a resonant value for the reverse waves under the condition of observa-
tion. In general, it is somewhat difficult to abstract power from a wave-
guide in a short length interval through a non-resonant slot even though
appreciable magnetic intensity is present to excite the slot. A preferred
form of this type of isolator might therefore be to replace the continuous
longitudinal slot with a series of slots having lengths which are resonant
in the frequency band of interest.

(c) A Field-Displacement Four-Terminal Circulator

A circulator utilizing the non-reciprocal field displacement idea is
sketched in Fig. 48. Due to the orientation of waveguide 2-3 relative to
waveguide 1-4, the only field component of waves propagating in guide
1-4 which will excite a wave in guide 2-3 is the longitudinal rf magnetic
intensity. The ferrite loaded guide 1-4 behaves in a manner very similar
to the description for the waveguide of Fig. 47. For propagation in the
direction of 4 to 1, the longitudinal h is zero to the left of the center
line of waveguide 1-4, and a series of coupling apertures between guides
2-3 and 1-4 are positioned along this null. As a result no power is coupled
from either guide into the other for waves propagating in this direction.
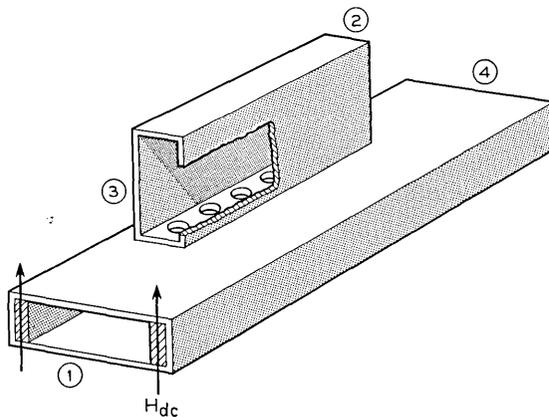
Fig. 48 — A field-displacement four-terminal circulator.

For propagation from 1 to 4, longitudinal h will appear at the apertures, and the number of coupling apertures and the coupling per individual aperture is predetermined,* so as to cause complete transfer of power from terminal 1 to terminal 2, and from terminal 3 to terminal 4. The direction of circulation is then $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1$.

(d)A Field-Displacement Three-Terminal Circulator

Application of the field-displacement effect to form a three-terminal circulator is sketched in Fig. 49. Guide 1-2 is once more ferrite-loaded and at the coupling slot which is common to branch guide 3 the longitudinal magnetic intensity is zero for one direction of transmission and finite for the reverse direction of transmission in guide 2-1. The coupling slot is so positioned that a wave entering terminal 1 does not excite the slot and therefore passes without reflection to terminal 2. This situation is sketched diagrammatically in Fig. 49(b) as transmission from terminal 1 to terminal 2 only. Energy entering at terminal 2 will propagate with a finite value of longitudinal rf magnetic intensity at the coupling slot and transmit power to both terminals 3 and 1. The energy which sets up a dipole at the longitudinal slot cannot radiate a reflected wave back toward terminal 2 since a wave propagating in the direction 1-2 has no magnetic intensity vector at the coupling slot. This situation is sketched
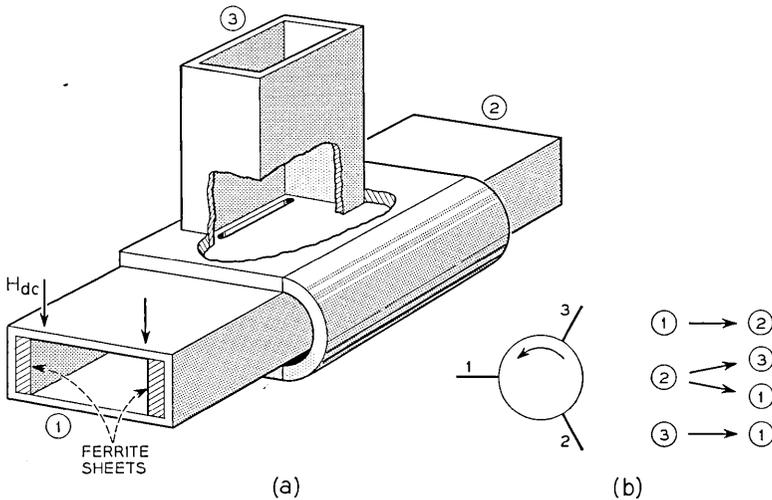
Fig. 49 — A field-displacement junction circulator.

* See S. E. Miller, Reference 8, pp. 694–695.

diagrammatically in Fig. 49(b) as transmission from terminal 2 to terminal 1 and to terminal 3. Energy which is introduced at terminal 3 will set up a magnetic intensity vector along the coupling slot which, for reasons already described, can radiate in the ferrite-loaded guide only in the direction of terminal 1. There may of course be a reflection component for a wave introduced at terminal 3 so that, in general, energy introduced at terminal 3 may propagate to terminal 1 or be reflected back upon itself.

If we now introduce in branch waveguide 3 a tuning reactance such as to cancel out the reflected wave for energy entering at terminal 3, we will transmit all of the power entering terminal 3 to terminal 1. This situation is sketched in Fig. 49(b) as transmission from terminal 3 to terminal 1 only.

The general set of conditions sketched in the power flow diagram of Fig. 49(b) can not exist when the external loads at terminals 1, 2 and 3 match the waveguide impedance. We may demonstrate this by observing that thermal noise power KTB flows from terminal 3 to terminal 1 but, as sketched in Fig. 49(b), there is also power flow from terminal 2 to terminal 1. This would seem to imply that more than KTB of energy is being transmitted to terminal 1, which is impossible. Actually, *under the conditions of matched input impedance at terminal 3* there can be no transmission from terminal 2 to terminal 1. Instead the transmission is from 1 to 2, 2 to 3, and 3 to 1, and the junction is a circulator.

One may visualize what has happened to the component of transmission from 2 to 1 as follows: the matching mechanism which was introduced in branch 3 to prevent reflection for waves entering at terminal 3 also reflects some of the wave energy which flows from 2 toward 3. This reflected component from the matching mechanism travels back toward
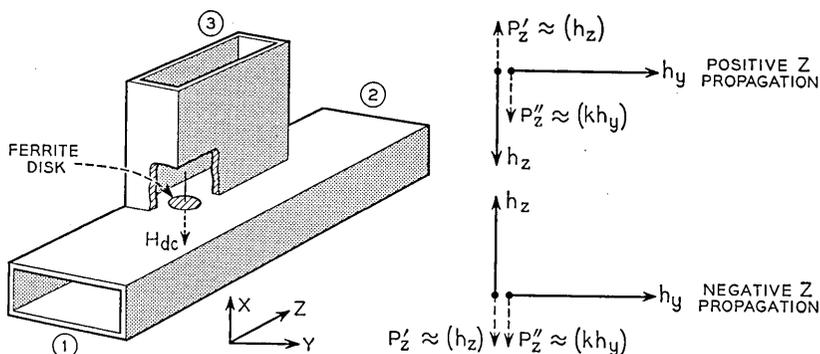
Fig. 50 — A junction circulator using a ferrite-loaded aperture.

the coupling slot and a portion proceeds in the direction of terminal 1 only. For the condition of matched input to terminal 3, the component of wave propagation $2 \rightarrow 3$ which is reflected from the matching mechanism will reappear in branch 1 in such an amplitude and phase as to cancel the direct transmission component of wave propagation $2 \rightarrow 1$.

Experimental models of the junction circulator of Fig. 49 have been built for operation at 6,000 mc and 24,000 mc, and the circulator type of transmission was observed. For the 6,000-mc model which used the same ferrite elements described above for the 6-kmc resistance-sheet isolator, the losses for the various transmission paths are tabulated below:

Path	Insertion Loss db
$1 \rightarrow 2$	0.7
$2 \rightarrow 3$	1.3
$3 \rightarrow 1$	1.3
$2 \rightarrow 1$	14.
$1 \rightarrow 3$	13.5

A rather lossy (sliding-screw) tuner was employed in branch 3 in this experiment, so these data do not represent the best that can be achieved. With a good piston in branch 3, the loss for path $2 \rightarrow 3 \rightarrow 1$ was observed to be about 0.1 db.

(e) *A 3-Terminal Junction Circulator Utilizing a Ferrite-Loaded Aperture*

E. H. Turner has pointed out that a ferrite-loaded aperture has non-reciprocal properties.* The present discussion indicates the way such an aperture can be utilized to form another kind of junction circulator. The structure is shown in Fig. 50. The coupling aperture, which is ferrite-loaded, is slightly to the left of the center-line of the guide 1-2, and filled with ferrite. The waveguides may otherwise be empty.

In order to explain the operation of this device, we note first that the branch guide 3 will be excited only by a Z-axis component of magnetic polarization in the aperture. Such a polarization will result from the component h_z in guide 1-2, and it will point in a direction opposite to h_z.† It is caused in part by the geometry of the aperture, and is further enhanced by the Z-axis component of ferrite magnetization due to h_z when $(\omega > \omega_0)$. In addition there is another Z-axis component of polari-

---

* Unpublished work.
† For a picture of magnetic polarization of an unloaded coupling aperture see R. L. Kyhl, Reference 8, p. 860.

zation caused by h_y as a result of the precession of ferrite magnetization about H_{dc} as explained in connection with Fig. 5.

Fig. 50 shows the time relations between the rf vectors for propagation in both directions in guide 1-2. h_y is the largest vector (since the aperture is near the centerline) and is taken as reference. The Z-axis polarization (p_z'') caused by h_y is proportional to kh_y, and lags h_y by 90 degrees. The Z-axis polarization (p_z') caused by h_z will have a phase which depends upon the direction of wave propagation in guide 1-2. For propagation in the plus Z direction, h_z will lag h_y by 90 degrees, and consequently the polarization will lead h_y by 90 degrees. It will be seen that the Z-axis components of polarization p_z' and p_z'' point in opposite directions. By adjusting the dc magnetic field (which controls k) or the displacement of the aperture from the centerline of the waveguide, the relative magnitudes of p_z' and p_z'' can be adjusted so as to obtain zero net polarization for this direction of propagation. Thus no wave will be excited in branch guide 3. For propagation in the minus Z direction, however, h_z will lead h_y by 90 degrees and the vector relations are given by the lower diagram of Fig. 50. In this case p_z' and p_z'' point in the same direction, and there will be a net polarization of the aperture.

Thus, for propagation $1 \rightarrow 2$, no transmission to 3 occurs, but for propagation $2 \rightarrow 1$ there will be transmission to 3. A line of reasoning similar to that given for Fig. 49 indicates that circulator behavior can also be obtained in this structure. In a 24,000-mc model of the structure of Fig. 50, transmission $3 \rightarrow 1$ was 14–20 db greater than transmission $3 \rightarrow 2$. No attempt was made to match the input impedance at terminal 3.

There is another induced component due to the ferrite response to driving force h_z which is either in phase or out of phase with p_y' (due to h_y) depending on direction of propagation. The guide 3 may be rotated 90° to respond to components of p in the y-direction only, and the circulator may be built in this form as an alternative.

6. MAGNETICALLY CONTROLLED RECIPROCAL PHASE SHIFT IN FERRITE-LOADED STRUCTURES

Although most of the attention of this paper has been directed at the non-reciprocal microwave properties of ferrites, their reciprocal properties, which are under the control of a magnetic field, can also be of great importance. Some aspects of these reciprocal effects have been discussed in an earlier publication[23] where it was shown that, for a uniform plane wave propagating through a ferrite magnetized perpendicular to the rf

magnetic vector, the rf permeability is given by:

$$\mu_\perp = \frac{\mu^2 - k^2}{\mu} = 2 \frac{\mu_+ \mu_-}{\mu_+ + \mu_-} = \left[\mu_0 + \frac{\gamma^2 M^2 - \gamma M \omega_0 \mu_0}{\mu_0(\omega_0^2 - \omega^2) - \gamma M \omega_0} \right] \quad (27)$$

independent of the direction of propagation. If the rf magnetic vector is parallel to the dc magnetization, the rf permeability no longer differs from μ_0. Thus, a transverse magnetic field causes the infinite ferrite medium to be reciprocally birefringent. This reciprocal effect can be large only in the general vicinity of resonance as may be seen by examining equation (27) and Fig. 6. Since μ_+ and μ_- have odd symmetry about $H_{dc} = 0$, the application of small magnetizing fields will cause nearly equal and opposite changes in the μ_+ and μ_-. To first order these effects cancel and as a result $\mu_\perp$ departs from μ_0 very slowly. For large values of field, the odd symmetry no longer holds and $\mu_\perp$ may be appreciably different from μ_0. Thus, for small values of applied magnetic field, (unsaturated region where M is proportional to H), $\mu_\perp - \mu_0$ varies as H_{dc}^2; for large frequencies $\mu_\perp - \mu_0$ varies as $1/\omega^2$ as seen from equation (27). In contrast, the birefringence for positive and negative circularly polarized waves due to an axial magnetic field is a first-order effect since it depends directly on μ_+ and μ_- whose departure from μ_0 is directly proportional to H_{dc}. Thus, $\mu_- - \mu_+$ is given by

$$\mu_- - \mu_+ = \frac{-2\omega\gamma M}{\omega^2 - \omega_0^2} \quad (28)$$

showing that this non-reciprocal effect is linear in M (rather than quadratic) and approximately proportional to $1/\omega$ (rather than $1/\omega^2$). Both $\mu_1 - \mu_0$ and $\mu_- - \mu_+$ are plotted in Fig. 51 as functions of ω/ω_0 for

$$-\frac{\gamma M}{\mu_0 \omega_0} = 1.5 \quad \text{and for} \quad -\frac{\gamma M}{\mu_0 \omega_0} = 0.5.$$

By reading the abscissa as a difference in "external μ's" (see end of Section 2) and the ordinates as the ratio of ω to the external resonant frequency, these curves become applicable to cases frequently encountered in practice. If we bias a given ferrite at a fixed H_{dc} sufficient to bring about saturation, then M and ω_0 are constant and we can use the plots of Fig. 51 to compare the reciprocal and the non-reciprocal effects at various signal frequencies, ω.

A comparison between reciprocal and non-reciprocal phase shifts for various positions of a longitudinally placed ferrite slab in a rectangular waveguide at 10,500 mc can be obtained from Fig. 52. The non-reciprocal phase shift plotted is the *difference* in phase for the two directions
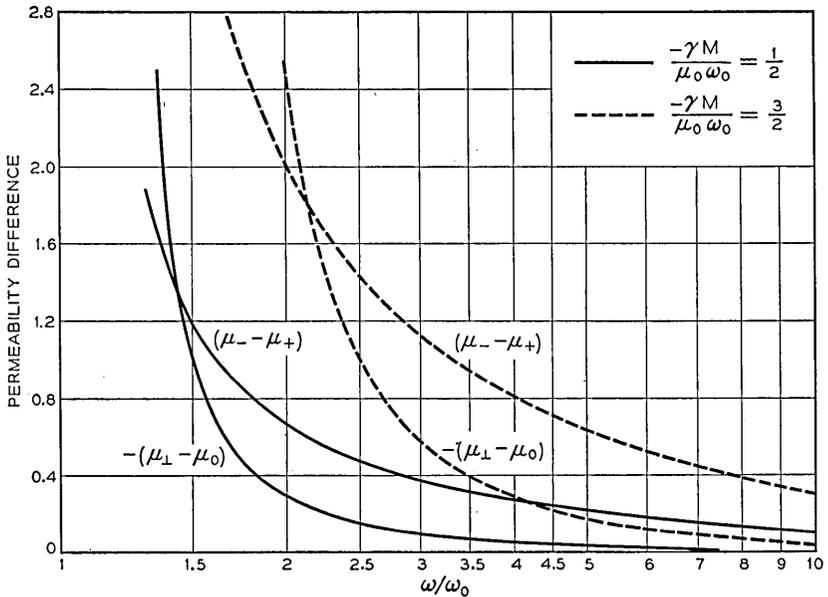
Fig. 51 — Calculated reciprocal $(\mu_{\perp} - \mu_0)$ and non-reciprocal $(\mu_- - \mu_+)$ ferrite permeability differences as a function of frequency.

of transmission for *fixed* magnetizing field, whereas the reciprocal phase plotted is one-half the sum of the phase changes for the two directions of propagation caused by going from zero magnetizing field to abscissa value of magnetizing field. With the ferrite at the wall of the waveguide, there is a small amount of reciprocal phase shift varying as H^2 with negligible non-reciprocal phase shift for fields below resonance. As the ferrite is moved away from the guide wall, the reciprocal phase shift increases because more energy is concentrated in the ferrite. The non-reciprocal phase shift increases very rapidly as discussed in Section 5-2. With the ferrite moved to the center of the guide, the non-reciprocal phase shift is reduced to zero with the reciprocal phase shift being at a maximum.

7. BIREFRINGENCE IN ROUND WAVEGUIDE

A round waveguide containing an axially symmetric ferrite element is inherently a two-mode guide. If the ferrite is unmagnetized, normal modes are degenerate in the sense that they have the same phase veloc-ity, and we are at liberty to define them either as two orthogonal linearly

polarized waves, or as (+) and (−) circularly polarized waves. With the application of any magnetic field, the degeneracy is removed. For longitudinal magnetization the circularly polarized waves remain the normal modes, while the linearly polarized waves become coupled modes. On the
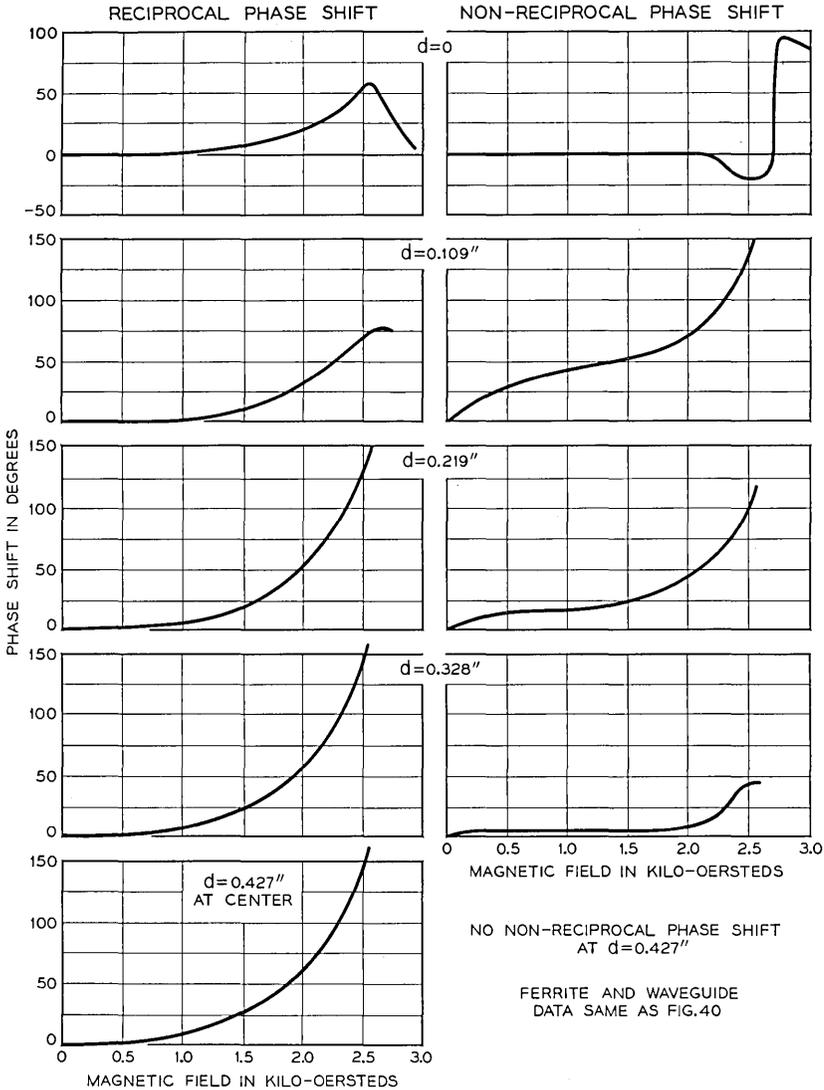
Fig. 52 — Measured reciprocal and non-reciprocal phase shifts for the ferrite-loaded rectangular guide of Fig. 40.

other hand, if the ferrite is magnetized in a transverse direction, the normal modes are two linearly polarized waves having different phase velocities, and the circularly polarized waves become the coupled modes. We may therefore note a form of duality: when longitudinally magnetized, a ferrite medium is birefringent for circularly polarized waves; and when transversely magnetized it is birefringent for linearly polarized waves. The former type of birefringence results in Faraday rotation. The latter type of birefringence has as yet received comparatively little attention but it is likely to prove important in a number of applications. It can be manifested as either a reciprocal or a non-reciprocal effect, and the relative importance of these two effects depends upon the frequency of the rf wave and upon the direction of the magnetization.

The reciprocal birefringence has been discussed for the plane-wave case in Section 6. We now consider this effect for a slender axial pencil in round waveguide, as shown in Fig. 53. With magnetization normal to the paper, a point-field examination indicates that excitation of the ferrite by the dominant wave is essentially transverse (dominant wave
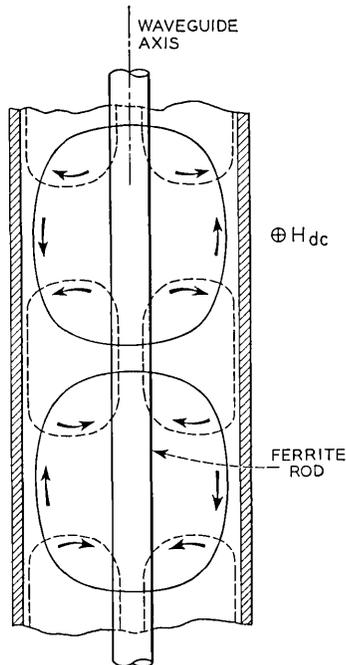
Fig. 53 — Longitudinal section of ferrite-loaded round waveguide with transverse magnetizing field.

magnetic loops shown solid). Ninety degrees later in time the precession of electron spins will produce an axial magnetization at every point where it was formerly transverse. This will tend to generate new b loops shown dashed. These correspond qualitatively to the magnetic field pattern of a TE_{01} wave. Thus, as a byproduct of our examination, we discover that this structure will tend to convert dominant waves into circular electric waves. However, if circular electric waves cannot be propagated, the dashed b loops cannot exist. Instead, an added reactance is presented to the driving wave similar to that which appeared in the infinite transversely-magnetized medium (Section 6). For the orthogonal TE_{11} wave (rf h parallel to the magnetizing field), μ equals μ_0 and the magnetizing field does not alter the phase constant. Thus, the medium is birefringent due to the transverse field, and this birefringence is small for frequencies far above the resonant frequency.

We may also look at the symmetrical ferrite in a somewhat different way. Assuming that the ferrite does not perturb the dominant mode field pattern, all ferrite lying to the right of the center line of Fig. 53 will be exicted by $(+)$ elliptically polarized rf magnetic field vectors. All ferrite to the left of the center line will be excited by $(-)$ eliptically polarized field vectors. As a result, ferrite to the right of the center line will exhibit a permeability less than μ_0, while ferrite to the left of the center line will exhibit a permeability greater than μ_0. To a first approximation, these two effects tend to cancel one another and the total effect on the phase velocity of the dominant wave will be small. We again conclude that the reciprocal birefringence which results from a symmetrical ferrite structure will be small unless the resonant frequency approaches the frequency of the applied wave.

As in the case of the rectangular waveguides considered earlier, this first order cancellation of $(+)$ and $(-)$ permeabilities can be eliminated
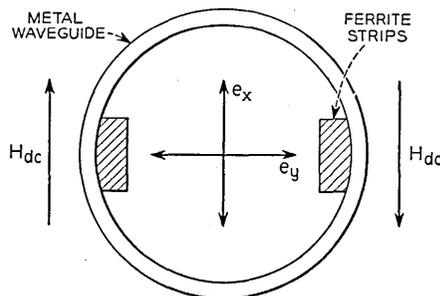
Fig. 54 — Cross section of round waveguide with ferrite next to the metal wall.

by reversing the dc magnetic field on opposite sides of the waveguide center line. Fig. 54 shows the cross section of a waveguide containing two strips of ferrite running along opposite walls and magnetized in opposite directions. If a dominant wave having electric polarization e_x propagates into the paper, the permeability of both strips of ferrite will appear to be less than μ_0, and there will be a first order effect on the phase velocity of this wave (aside from any dielectric effect). A dominant wave having polarization e_y should see a permeability of μ_0 since the rf magnetic vector is parallel with the dc magnetization, and there will be no effect on the phase velocity of this wave (aside from the di-
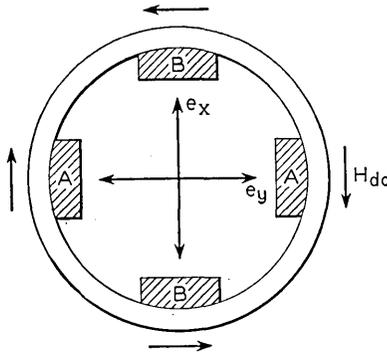
Fig. 55 — Further development of structure of Fig. 54 to eliminate dielectric birefringence and enhance the magnetic birefringence.

electric effect). Thus, there will be a magnetically induced birefringence for waves of these two polarizations.

In addition to the magnetically induced birefringence, the structure of Fig. 54 also introduces a dielectric birefringence due to the presence of the ferrite. This dielectric birefringence can be eliminated and the magnetic birefringence augmented through the use of the structure show in Fig. 55. For polarization e_x propagating into the paper, ferrite strips A-A produce a permeability less than μ_0, while ferrite strips B-B have no magnetic effect on this wave. For polarization e_y propagating into the paper, ferrite strips B-B have a permeability greater than μ_0, while strips A-A have no effect. Consequently, polarization e_x will have a higher phase velocity and polarization e_y will have a lower phase velocity than if the dc magnetic fields were not present. If the direction of propagation is reversed, and waves travel out of the paper, polarization e_x will have the lower phase velocity and polarization e_y will have the higher phase velocity. As will be shown presently, such a non-reciprocal

birefringent medium can be made to perform all of the functions of a Faraday rotation medium, and with it circulators, isolators, and gyrators can be built.

The arrangement of Fig. 55 is not the only one which exhibits strong birefringence. Another arrangement is shown in Figure 56. A vertically polarized wave e_x will produce rf magnetic vectors next to the wall of the waveguide which are transverse at points 1 and 3 and longitudinal at points 2 and 4. In between 1 and 2, as for example at A, the magnetic vector will be elliptically polarized tangential to the wall. If a piece of ferrite is placed at A and a dc magnetic field is oriented normal to the wall as shown, a wave e_x propagating into the paper will see a permeability less than μ_0 in the ferrite, and a wave e_y will see a permeability greater than μ_0. In this manner E. H. Turner obtained the first non-reciprocal birefringent medium.[24] While the single ferrite strip has both magnetic and electric birefringence, the addition of the other three ferrite strips, B, C and D of Fig. 56, eliminates dielectric birefringence. With dc field applied as shown, all four strips have a permeability less than μ_0 for e_x propagating into the paper, and a permeability greater than μ_0 for e_y propagating into the paper. For the opposite direction of propagation, the phase velocities are interchanged, and this medium is also non-reciprocally birefringent.

It is possible to interlace the structures of Figs. 55 and 56 to obtain the structure of Fig. 57. With external magnetic poles shown, the magnetization of the ferrite ring will be such as to make all parts of the ferrite active in producing birefringence. The section shown in Fig. 58 was built to test this principle at 24 kmc. It was verified that while a simple
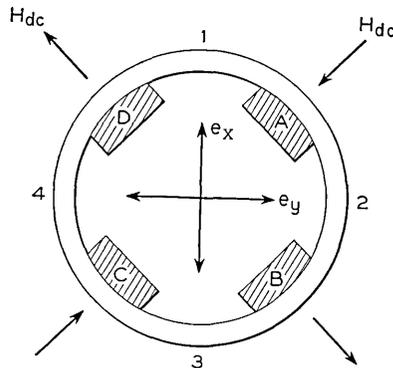
Fig. 56 — Alternate form (to that of Fig. 55) for a magnetically birefringent structure.

transverse magnetizing field produced no noticeable birefringence, a
4-pole field as shown in Fig. 57 produced a differential phase shift of 180°.
Thus, as the 4-pole field was rotated a given amount, the output polari-
zation was rotated by twice this amount. This technique may be useful
in constructing continuously running phase changers, as will be shown
later.

The shape and positioning of the ferrite in the above structure was a
compromise selected for its symmetry and mechanical simplicity. Clearly,
the ferrite is not used in the most effective places in the waveguide, since
where the dc magnetization is tangential to the wall, the rf field is almost
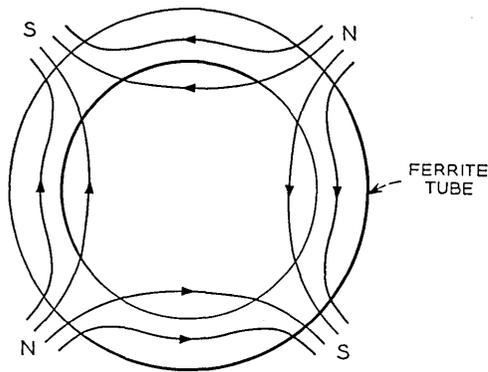
Fig. 57 — Composite of Figs. 55 and 56.

linearly polarized. We may then ask where the ferrite should be placed
in order to be most effective.

In order to orient ourselves with regard to the field components in
round guide we may note that the total h vector is circularly polarized
at the loci drawn in Fig. 59, which also shows for comparison the loci of
circularly polarized h in a rectangular guide having the same cut-off
frequency. In the round guide it is important to note that the loci are
normal to the planes of the polarization vectors at all points, and the
dc magnetic field should be directed along the loci for maximum effec-
tiveness.

As in rectangular waveguide, the position of pure circularly polarized
h is *not* the optimum ferrite location for producing non-reciprocal phase
shift. The latter locus is along points for which the product of trans-
verse-h and longitudinal-b is a maximum, as discussed in Section 5-2.
Fig. 60 illustrates such a locus for an unperturbed guide (i.e., for thin
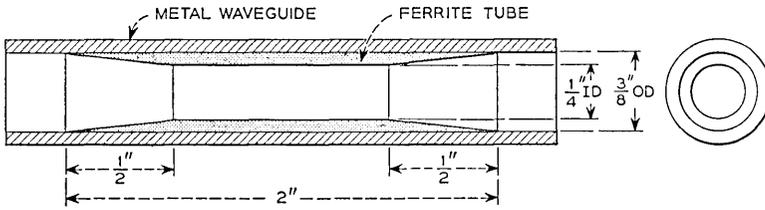
Fig. 58 — Longitudinal section showing the dimensions of a 24-kmc model of the structure of Fig. 57.

ferrite sheets added). Just as in rectangular guide, the most effective ferrite location along axis b–b is about $\frac{1}{4}$ the diameter away from the wall. The optimum ferrite location near the wall can be obtained by noting that h is trnsverse at points a and longitudinal at points b, with sinusoidal variation between. Therefore, the product of h transverse and h longitudinal will be maximum half way between points a and b, namely,
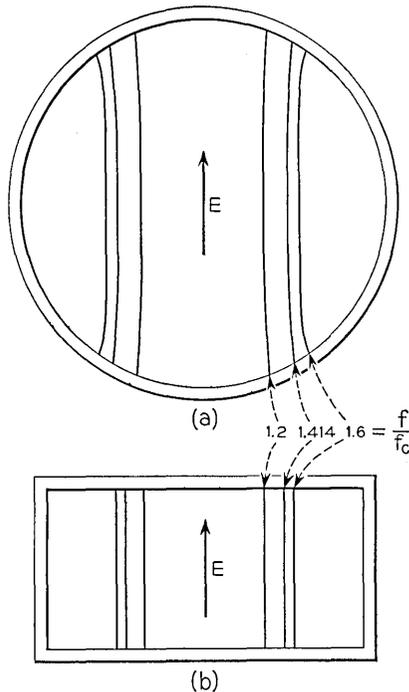
Fig. 59 — Locus of points of circularly polarized magnetic intensity for $TE_{11}\bigcirc$ and $TE_{10}\square$ in guides having the same cut-off frequency.

at points 2. Thus, there will be a locus of points shown by the dashed lines, along which a tiny amount of ferrite will be most effective in perturbing the phase velocity of a vertically polarized wave, and these loci are independent of frequency.

As in the rectangular waveguide case, when the amount of ferrite present is sufficient to perturb the wave, the position for maximum effectiveness moves toward the side wall, as illustrated in Fig. 61 for both TE_{11} waves. The thin-walled cylinder of Fig. 58 is a practical approximation to this optimum.

7-1. *Equivalence of Faraday Rotation Elements and Non-Reciprocal Birefringent Elements*

Linear polarization birefringent elements of the type described are quite different from Faraday rotation elements in the way they transform the polarization of transmitted waves. Yet both are basically two-mode elements, and either one can be made to duplicate the performance of the other. Thus, by combining two non-reciprocal birefringent elements with a reciprocal birefringent element, a Faraday rotation element can be simulated. This is indicated in Fig. 62.

Elements a and c are reciprocal $\Delta 90°$ birefringent elements having their axes of low phase velocity oriented vertical and horizontal as indicated by the solid lines. Element b is a non-reciprocal birefringent element having θ degrees differential phase delay with its axes of low phase velocity oriented at 45° to the vertical as shown by the solid and
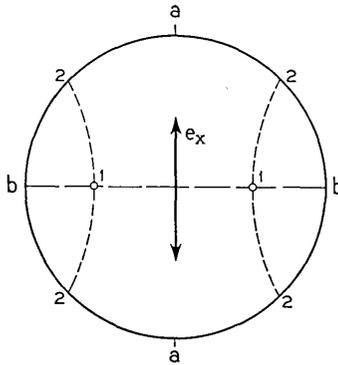
Fig. 60 — Calculated locus of optimum ferrite locations for producing non-reciprocal phase shift of the $TE_{11}\bigcirc$ wave (applicable only for very thin ferrite sheets).

dashed lines for propagation from left to right and right to left respectively. Similarly, a Faraday rotation element combined with two reciprocal birefringent elements is the equivalent of a non-reciprocal birefringent element as shown in Fig. 63.

In practical applications some of the reciprocal elements may be superfluous, and simpler equivalent structures can be used. For example, a Faraday rotation type of circulator can be built using one non-reciprocal and one reciprocal birefringent element as shown in Fig. 64. Element a may contain a ferrite tube with a four-pole transverse magnetic field producing a differential phase shift of $90°$. Its axis of high refractive index is at $45°$, as shown by the solid line, for propagation from left to
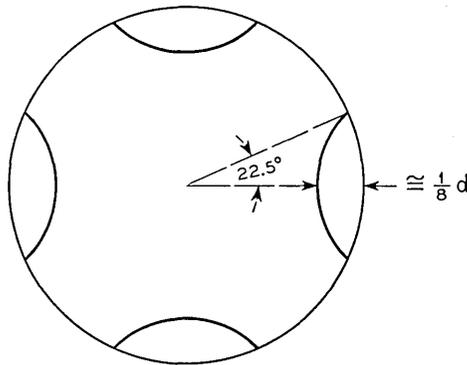
Fig. 61 — Estimated optimum positions of thick ferrite elements for producing non-reciprocal phase shift of the $TE_{11}O$ wave.

right. Element b is a reciprocal $\Delta 90°$ phase shift element with its axis of high refractive index at $45°$ as shown. For propagation from left to right the two sections constitute a $\Delta 180°$ section and polarizations 1 and 3 are rotated into polarizations 2 and 4, respectively. For propagation from right to left the axis of high refractive index for section a is the dashed line. Thus, the two sections constitute a $\Delta 0°$ element and the polarizations are not rotated; polarization 2 is transmitted to polarization 3, and polarization 4 is transmitted to polarization 1. Thus, these two elements comprise a circulator with commutation in the sequence of numbering. This combination is more flexible than the equivalent $45°$ Faraday rotation circulator in that the output polarizations may be arbitrarily oriented with respect to the input polarizations by orienting the axes of section b relative to the axes of section a.

7-2. *Other Transverse-Field Effects in Round Waveguide*

By using circumferentially magnetized ferrite, it is possible to obtain non-reciprocal phase shift, loss, or field displacement for any transverse electric mode. When a thin-walled cylinder of ferrite is mounted coaxially in a round waveguide and magnetized in a circumferential direction, the components of the rf magnetic field transverse to this magnetization are elliptically polarized in the same sense in all parts of the ferrite. The ferrite will produce a non-reciprocal phase delay for any transverse-electric wave, and because of the symmetry of the system, there will be no birefringence. By requiring that the radial h be equal to the longitudinal h, we may determine the locus of points for which the component of the magnetic vector perpendicular to the dc magnetization is circularly polarized. This locus is defined by the following equation for TE_{nm} waves:

$$\frac{J_n'\left(k_{nm}'\,\frac{\rho}{R}\right)}{J_n\left(k_{nm}'\,\frac{\rho}{R}\right)} = \frac{1}{\sqrt{\left(\frac{2\pi R}{\lambda_0 k_{nm}'}\right)^2 - 1}} \tag{29}$$

where R = radius of the waveguide
λ_0 = free space wavelength
ρ = radial distance coordinate inside waveguide
J_n and J_n' = Bessel function of nth order and its derivative, respectively
k_{nm}' = m^{th} positive root of $J_n'(x) = 0$ which is 1.84 for TE_{11}

Equation (29) may be solved for the fractional radius ρ/R, and we find that the locus of circular polarization relative to a circumferential
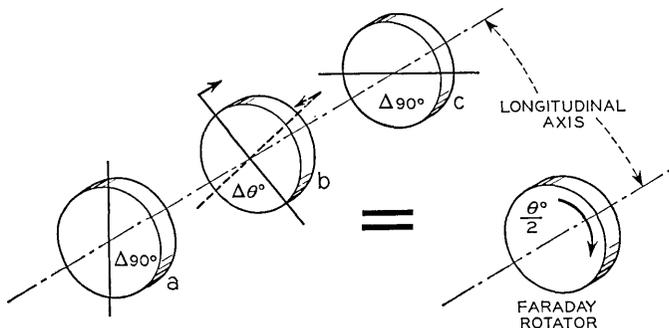
Fig. 62 — Diagram of a birefringent structure which is electrically equivalent to a Faraday rotator.

magnetization will be a circle whose radius depends upon the proximity of the operating frequency to cut-off frequency, as shown in Fig. 65. Again we note, however, that the locus of positions for maximum effectiveness of a given amount of ferrite in producing non-reciprocal phase shift will not be the same as the locus for pure circular polarization. This locus will also be a circle but will occur at a radius which produces a maximum value for the product of h longitudinal and h radial. For TE_{11} waves this occurs at a radius equal to 0.493 of the waveguide radius, and is independent of proximity to cut-off frequency.

It is noted above that the circumferentially-magnetized thin-walled cylinders of ferrite can be used to produce non-reciprocal phase delay
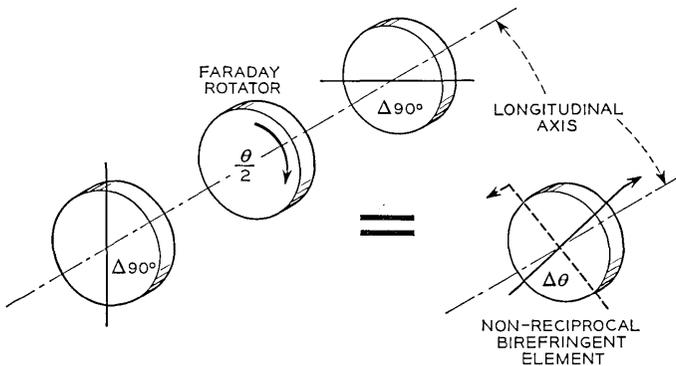
Fig. 63 — Alternate form showing equivalence between Faraday rotation and birefringent structures.

for any of the transverse electric modes of propagation. The optimum radius for the cylinder will, of course, depend upon which mode is used. For the circular electric mode (TE_{01}^O), there will be two radii at which the magnetic field vector will have pure circular polarization. At one of these the circular polarization will have one sense, and at the other it will have the opposite sense. In this case it would be possible to employ two thin-walled cylinders of ferrite placed at the two most effective radii, and magnetized circumferentially in opposite directions.

Since there is no angular variation in the h components for the circular electric wave, circumferential ferrite biasing for TE_{01}^O is the direct analogue of the transverse ferrite biasing for TE_{10} in rectangular guide. A field-displacement isolator analogous to Fig. 42 can be made for TE_{01}^O using two circumferentially magnetized ferrite cylinders, one coated with a resistance film.

8. NON-RECIPROCAL DIELECTRIC WAVEGUIDE CIRCUITS

The behavior of ferrite-loaded hollow metallic waveguides, as previously discussed, depends on wave-electron interactions which can also be utilized in other types of microwave transmission lines. For example, a number of the non-reciprocal transmission elements may be built in dielectric waveguide form.* Fig. 66 shows a longitudinal section of a round dielectric waveguide containing ferrite loading for the purpose of producing Faraday rotation. The rf magnetic intensity vector at the center of the dielectric waveguide is transverse and linearly polarized for a linearly-polarized lowest-order wave. Therefore, with longitudinal
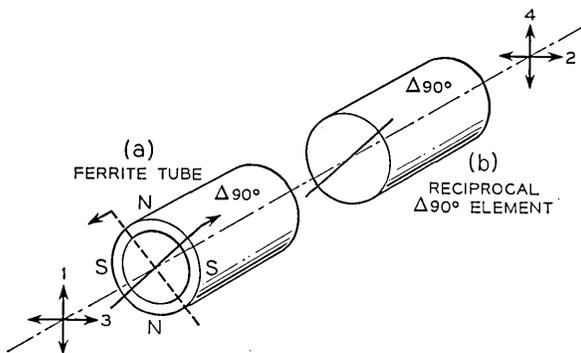
Fig. 64 — Birefringent elements capable of replacing a 45° Faraday rotator in a circulator.

magnetization, a linearly polarized input wave is coupled to the similar wave polarized at 90° to the first in a manner entirely analogous to the coupling between the two $TE_{11}{}^{\circ}$ waves in hollow metallic waveguide. Experiments were conducted at 48,000 mc using a structure of the form of Fig. 66. The ferrite diameter was $\frac{1}{8}''$ and the outside diameter of the polystyrene in the ferrite-loaded region was $\frac{1}{4}''$. The ferrite consisted of powdered Ferramic J molded in a polystyrene binder. More than 90° rotation was observed with a biasing field of a few hundred gauss; the transmission loss was $\frac{1}{2}$ db or less.

The Faraday rotator of Fig. 66 may be combined with dual polarization take-offs to form a circulator in dielectric waveguide as illustrated in Fig. 67. The dual-polarization take-offs have the function of providing

---

* The general principles involved in dielectric transmission lines are discussed in Reference 12, pp. 129–131. A. G. Fox gave a paper on dielectric-waveguide circuit elements at the March, 1952, I.R.E. National Convention.

separate terminals for the two polarizations of dominant wave at both ends of the Faraday rotator. One polarization of wave on the round dielectric rod passes by the region in which it is coupled to the rectangular dielectric guide with negligible loss, because the phase constants of the dielectric waveguide and rectangular waveguide are dissimilar.* The other polarization of wave on the round dielectric guide is arranged to have the same phase constant as has the parallel polarization on the rectangular guide, and therefore conditions may be so adjusted as to produce complete power transfer between the two guides. On a 48,000 mc experimental model of such a dual polarization take-off the transfer loss for the polarizations having the same phase constants was about ½ db, and the straight-through loss on the round rod for the polarization at 90° to the above waves was 0.1 db. In Fig. 67, a wave entering at 1 is transferred completely to the parallel polarization of wave on the round rod at the input to the Faraday rotator, and after rotation emerges in the same polarization as the take-off labeled 2. Transmission in at 2 again produces a wave of corresponding polarization on the round rod, which is then rotated an additional 45° and emerges on the near end of the Faraday rotator in the polarization labeled 3 which passes by the coupling region unaffected. In similar fashion, transmission occurs from terminal 3 to terminal 4 and from terminal 4 to terminal 1.

Non-reciprocal phase shifts in ferrite-loaded rectangular dielectric
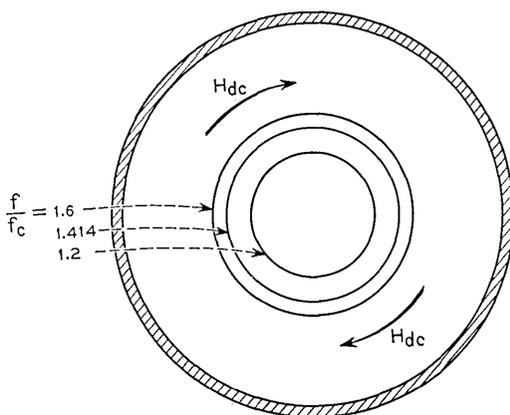
Fig. 65 — Calculated locus for which the component of h perpendicular to the circumferential dc magnetization is circularly polarized.

* This type of behavior is discussed in more detail in pp. 680–682 of S. E. Miller, Reference 8.
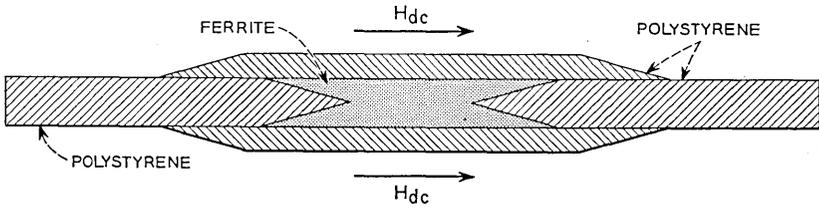
Fig. 66 — Faraday rotator for dielectric waveguide.

guides may be obtained in a manner entirely analogous to that described for the rectangular hollow metallic waveguide. Transmission line 3-2 of Fig. 68 contains transversely biased ferrite asymmetrically positioned on the dielectric guide. The rf magnetic intensity at the ferrite is elliptically polarized with respect to an observer viewing along the biasing field, and the sense of the predominant circularly polarized component reverses when the direction of propagation is reversed. Thus, the phase constant is non-reciprocal. Non-reciprocal phase shifts have been observed on this structure with negligible transmission loss, but no extensive data were taken.
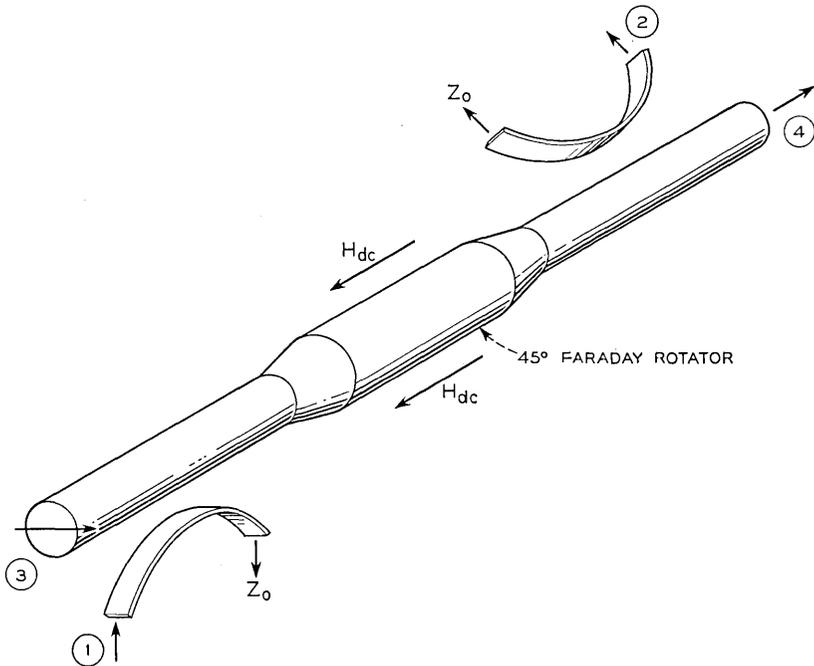
Fig. 67 — Dielectric waveguide circulator using Faraday rotation.

A coupled wave circulator in dielectric waveguide is represented by the entire sketch of Fig. 68, and its principle of operation is the same as discussed in Section 5-4 in connection with Fig. 36.

9. APPLICATIONS

9-1. *Isolators*

From the standpoint of its wide applicability, perhaps the isolator would be judged the most useful of the non-reciprocal ferrite devices When located next to the oscillator in a long waveguide connecting the oscillator to an antenna, the frequency pulling of the oscillator by the "long line effect" can be substantially reduced. This greatly improves the linearity of frequency modulation which can be obtained with klystrons.[18] If the impedance of the isolator as seen from the antenna side is a better match than the oscillator itself, then echoes will be reduced which could otherwise be multiply reflected between the oscillator and the antenna before finally being launched by the antenna. Further reduction of these echoes can be obtained by inserting next to the antenna
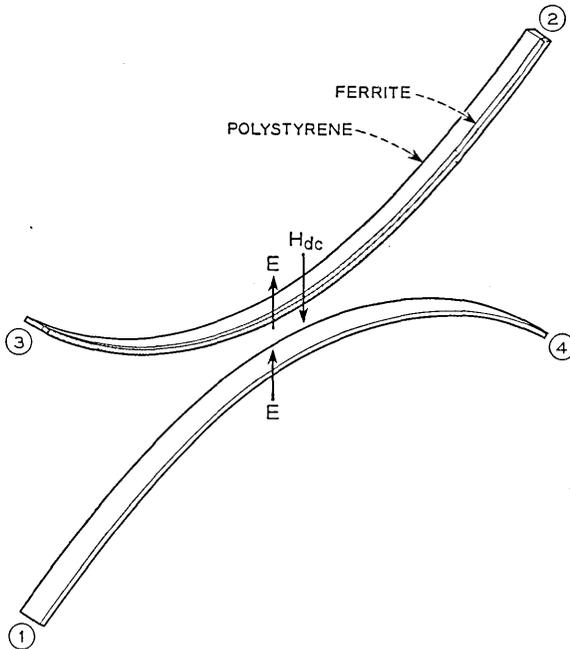
Fig. 68 — Dielectric waveguide circulator depending on non-reciprocal phase constants.

a second isolator having an input impedance match which is appreciably better than that of the antenna itself. For such systems applications of isolators, the requirements on the isolator for impedance match and discrimination against backward travelling waves are frequently rather stringent. Isolators can also be of very great value in experimental and measuring techniques. Here the requirements may be very much less stringent. Even a transmission ratio as low as 9 to 1 in decibels can be of great use. Thus, in most cases where a 10 db attenuator pad is used in a measuring setup, it could be replaced by an isolator having a forward loss of 2 db and a reverse loss of 18 db. The round trip attenuation will be the same in either case, but there will be an 8 db increase in the available forward traveling wave. Where the available signal power is low, and it is important to have a large measuring range, several such substitutions can increase the measuring range by 20 to 30 db.

9-2. *Circulators*

Circulators are considerably more complicated circuit elements, and the uses proposed for them are typically more sophisticated than those for isolators. An important distinction between the circulator and the isolator is that the circulator diverts or otherwise makes use of power which is reflected from the receiving end of the system rather than destroying it, as in the case of an isolator. One possible application of circulators is in channel frequency filtering systems. Such a system requires two vital components. It requires a filter which can pass certain frequencies and reflect others; and it requires some power dividing network which can separate out the reflected frequencies for further use. In other words, it is normally not satisfactory to allow the frequencies rejected by the filter to return toward the source. One method of constant impedance branching which can separate channels from one another without reflecting energy toward the source has been proposed by one of the authors, as shown in Fig. 69.[25] Incoming wave power containing all channel frequencies is divided at a first hybrid junction and passes to two identical transmission filters tuned for one of the channels. The channel which passes through these filters is recombined at a second hybrid and drops out at the point marked Channel 1. All other channels are reflected from the two pass filters but, upon returning to the first hybrid junction, differ in their round trip delay by 180° by virtue of a $\lambda/4$ difference in path length between the two filters and the hybrid junctions. As a result, all the rejected channels emerge from the remaining arm of the first hybrid junction and pass on down into the second such

network. A dual of this scheme has been proposed by W. D. Lewis.[26] The structure of a single channel-dropping section is the same as that of Fig. 69, but in this case the filter pairs are of the band rejection rather than the band pass type. All frequencies *except* the desired channel pass through the filters and are recombined at the second hybrid to pass on to a second such dividing network. These examples of constant impedance branching filters are given to indicate the degree of complexity required and to show how circulators can simplify this complexity appreciably.

Fig. 70(a) is a diagram showing how circulators can be combined with pass filters to provide a constant impedance channel-branching system. All channel frequencies entering from the left pass around the first circulator and impinge on the pass filter for channel 1. Channel 1 is dropped out at this filter and all other channel frequencies are reflected back to the first circulator, where they pass on around to the channel 2 filter. This process continues indefinitely with as many circulators and filters added as are necessary to separate out the channels. Nowhere is power reflected back to the source. Any residual power at the end of the array may be absorbed in a resistive load. It can be seen that a pair of hybrid junctions and a pair of identical matched filters for each channel
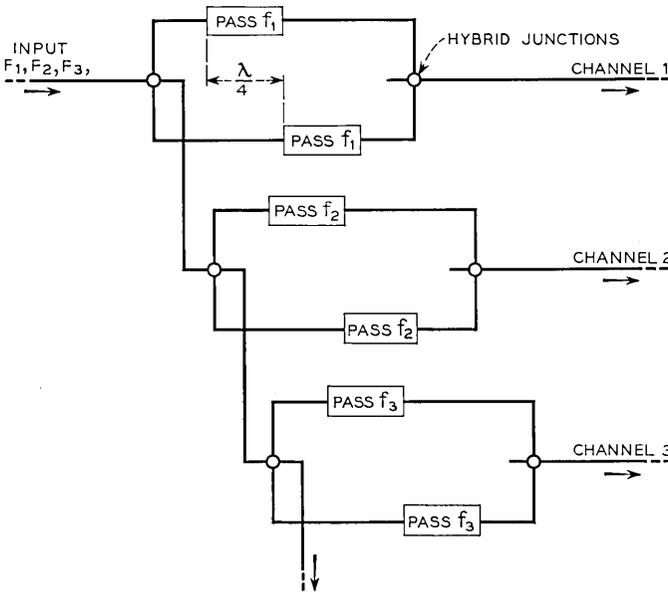
Fig. 69 — Channel-branching filter system using conventional hybrids.

in Fig. 69 have been replaced by a single transmission filter and one-half a circulator in the configuration of Fig. 70(a).

In Fig. 70(b) is shown an arrangement similar to the one of Fig. 70(a), except that rejection filters have been used. In this example only three arms are needed for the circulator. If four arms are available, the fourth one may be terminated in a characteristic impedance.

The three-arm circulator of Fig. 49 may be particularly well suited for use in filter applications, where channels are to be dropped off through pass filters. The tuning required to make such a circulator function can be provided as part of a transmission filter occupying the side arm of the device. As a result, a particularly convenient mechanical layout can be achieved where the main line proceeds straight through the structure, as shown in Fig. 71, with each of the channels dropping out of one of the side arms which project from the main line structure.

9-3. *Phase Changers*

A variable microwave phase changer described by one of the authors[27] has proven to be of considerable value in antenna scanning arrays[27] and in certain microwave measuring techniques.[28, 29] In these applications,
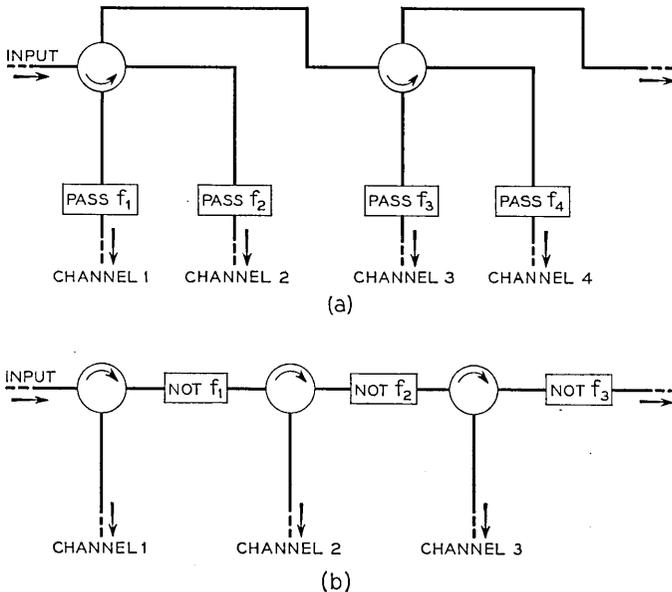
Fig. 70 — Channel branching with circulators.
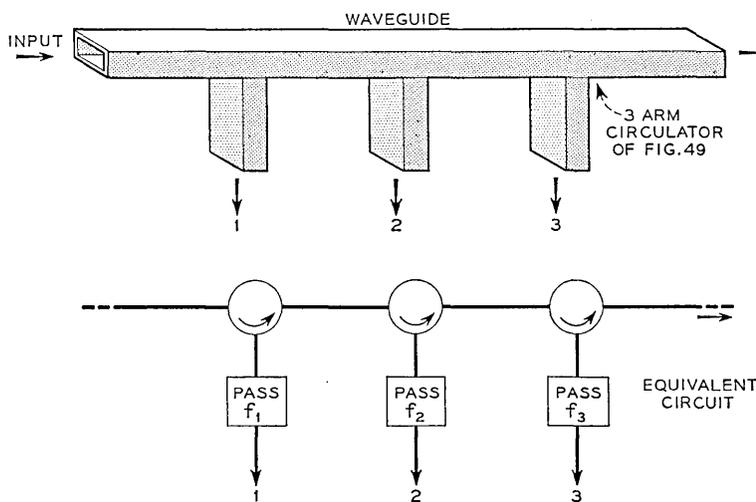
Fig. 71 — Waveguide embodiment of a channel branching filter using three-arm junction-circulators.

the 180° differential phase section in the middle of the device [Fig. 72(a)] is mechanically rotated in order to change overall phase shift. In some instances it would be very much more convenient if this rotation could be made to take place electrically. Since the function of this Δ180° section was to rotate the output polarization relative to the input polarization of this section, we can substitute a rotary medium comprising an axial pencil of ferrite in a round section of waveguide, so as to obtain the structure of Fig. 72(b). By varying the applied longitudinal magnetic field, the ferrite effectively rotates the output polarization relative to the input polarization of the middle section and achieves the same effect as physical rotation of the Δ180° section of Fig. 72(a). If a limited amount of phase shift (less than 1,000°, for example) is required, this Faraday-rotation type of phase changer can provide a convenient way of obtaining variable phase shift by means of electrical control. The limitation of the device is that the phase shift can be increased only up to the point where the ferrite is saturated. Beyond that point further increase of the applied field cannot change the output polarization, and as a result the total phase shift will be limited.

   In some of the applications of the mechanical phase changer a continuously rotating Δ180° section was necessary in order to obtain continuous change in phase. In order to apply ferrites to such a device, it would be necessary to employ a birefringent medium whose axes of

birefringence could be controlled magnetically. At frequencies of 10 kmc and below, the use of a round waveguide containing a rod of ferrite or a thin-walled tube of ferrite saturated by a simple transverse magnetic field, can produce enough birefringence to make such a Δ180° section possible. In this case the external magnet structure would be arranged like a rotating two-pole motor stator, and when properly excited, the axes of birefringence of the section could be made to rotate continuously. However, as pointed out earlier, this reciprocal birefringence may become so small at frequencies above 20 kmc as to be of little practical value. In this case, use may be made of the non-reciprocal birefringence obtainable from the four-pole field structures as described in Section 7.
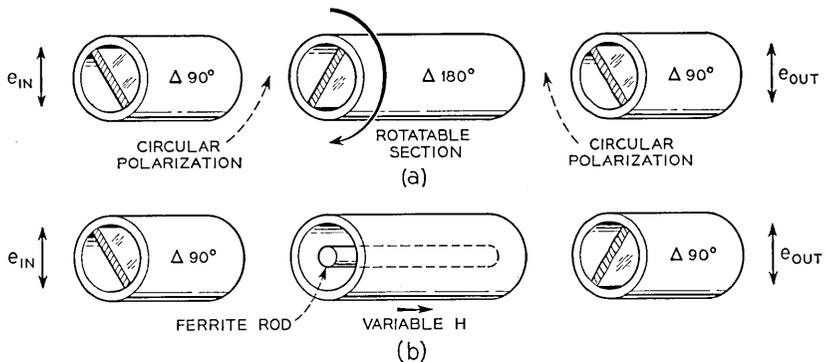
Fig. 72 — Phase changers — (a) using mechanical rotation, and (b) using variable magnetization of a ferrite element.

It is interesting to note that for pure non-reciprocal phase changers (which can be approximated using either the Faraday rotation element of Fig. 72(b) or the four-pole field birefringent element) no round-trip variation in phase delay is exhibited as the one-way phase delay is changed. In other words, the phase change for the waves traveling in one direction is equal and opposite to the phase change for waves traveling in the opposite direction. As a result, reflections arising on the load side of the phase changer will return to the source side with a phase (relative to the incident wave) which is independent of the phase changer setting. This is to be contrasted with the reciprocal type of phase changer for which the round trip phase delay is twice that of a one-way transmission.

10. SUMMARY AND CONCLUSIONS

The behavior of many configurations of ferrite-loaded waveguides has been deduced using a "point field" analysis. In this method it is assumed

that the ferrite-loaded waveguide contains a field configuration bearing a strong resemblance to one or more of the modes which are characteristic of the empty waveguide. The interaction between the added ferrite and the presumed distribution of field components is determined qualitatively at each point in the cross section, and the total behavior deduced as the weighted sum of the effects throughout the cross section.

The particular characteristic of ferrite materials which is responsible for the non-reciprocal behavior of ferrite-loaded microwave circuits is a tensor permeability — i.e., a driving-wave magnetic intensity in one direction may cause a component of magnetic flux at right angles to this direction as well as a component in the same direction. The magnetization of all "magnetic" materials may respond in the above manner, but the coupling between the electrons responsible for this behavior and the microwave field components has been found useful only in ferrites, which have negligible conductivity and within which microwave energy can propagate without appreciable attenuation.

Section 3 explains the effects of ferrite-loading a two-mode transmission medium (such as a dominant-wave round guide) using two viewpoints: (1) the normal mode approach in which the input and output waves are resolved into two modes which are orthogonal in the *presence* of the ferrite, and (2) the coupled-mode approach in which the behavior of the ferrite-loaded region is described in terms of a coupling introduced by the ferrite between the two modes which are orthogonal in the *absence* of the biased ferrite. These concepts are applied to Faraday rotation in Section 3, and to birefringence in ferrite-loaded round waveguide in Section 7.

Section 4 contains a discussion of the principles of Faraday rotation elements. The optimum location for the ferrite element in Faraday rotators is shown to be on the waveguide axis, and it is pointed out that there is an optimum diameter of ferrite rod with respect to obtaining a maximum of Faraday rotation per unit transmission loss. The problems encountered in impedance matching at the ferrite ends are discussed. It is shown that mode-conversion difficulties can appear as impedance match or Faraday rotation irregularities. The reasons for the variation of Faraday rotation as a function of frequency are given, and the dependence of rotation and transmission loss on temperature and on the Curie point of the ferrite is also discussed.

An important use of ferrite-loaded waveguide circuits lies in building passive non-reciprocal devices. It is pointed out that there are numerous forms of non-reciprocal devices which accomplish the same functions, just as there are numerous forms of amplifiers and filters. The authors

propose the names for the new elements be based on function and, in accordance with this point of view, definitions are given for the gyrator, circulator, and the isolator (Section 1).

Methods of employing Faraday rotation for the construction of gyrators, isolators, circulators and non-reciprocal power dividers are described, and a novel method of avoiding the deterioration of isolator performance due to frequency or temperature changes is described (Fig. 18).

A rectangular waveguide suitably loaded with ferrite can be non-reciprocal despite the existence of only one mode of propagation in each direction. Such guides may be non-reciprocal with respect to (1) phase constant (2) attenuation constant (3) magnetic and electric field configurations in the cross section. The circuits using these effects are fundamentally different from those depending on Faraday rotation, since the latter depend on the existence of two orthogonal waves in the ferrite-loaded region. The point-field concept is utilized to explain the non-reciprocal properties noted above and is used to indicate approximately the optimum ferrite position and direction of magnetization associated with obtaining these effects. When developing final models, one utilizes the maximum amount of ferrite which is tolerable in order to maximize these non-reciprocal effects, and it is indicated that the appearance of higher-order modes of propagation frequently limits the amount of ferrite which can be so utilized. Calculations are presented to serve as guidance for avoiding higher-order mode propagation, but the configuration of ferrite loading which is most attractive from the practical point of view is one which has not yet been adequately analyzed theoretically. Rectangular-waveguide isolators, circulators, and directional phase shifters are described, and experimental data on the performance of many of them are reported. In particular, the *resonance isolator* (Figs. 39–41, incl.) and the resistance-sheet type of *field-displacement isolator* (Figs. 42–46, incl.) have been explored in some detail.

Reciprocal phase shifts in ferrite-loaded structures may be important since they are under the control of a magnetizing field which may, under certain circumstances, be varied rapidly. In Section 6, the particular ferrite characteristics responsible for reciprocal phase shifts are discussed and the reciprocal phase shift obtained in practical ferrite-loaded structures is compared with the non-reciprocal phase shift in a similar structure. It is shown that the reciprocal phase shift requires ferrites of such saturation magnetization and with such magnetizing fields that the tensor permeability component k^2 is appreciable compared to tensor permeability component μ^2 (terminology defined in Section 2). This

implies operating the ferrite in a region not too remote from ferromagnetic resonance, but despite this condition very usable reciprocal phase shifts can be obtained with negligible ferrite loss. Reciprocal phase changes vary as the square of the applied field, whereas non-reciprocal phase changes vary linearly with the applied field.

Section 7 discusses birefringence in round waveguide. It is pointed out that the orthogonal TE_{11} waves for round guide containing an axially symmetric ferrite element are no longer degenerate when a magnetizing field is applied in any direction. For transverse magnetization, the medium is birefringent for the linearly polarized TE_{11} waves. Such birefringence may appear as either a reciprocal or non-reciprocal effect, the relative importance being determined by the frequency of the rf wave and by the detailed shaping of the magnetization in the transverse plane.

Section 7-1 compares the operation of non-reciprocal birefringent elements with Faraday rotation elements, and shows how either type can be substituted for the other in building non-reciprocal circuits.

Section 7-2 shows the manner in which non-reciprocal phase shift, loss, or field displacement may be obtained for any of the transverse-electric modes in a round waveguide through the addition of transversely magnetized ferrite. The optimum conditions for obtaining non-reciprocal phase shift or non-reciprocal loss are discussed. Faraday rotation for any of the transverse electric modes in round or square waveguide can be obtained, and the optimum geometry of ferrite for producing such rotation is indicated.

The wave-electron interactions which have been utilized in ferrite-loaded hollow metallic waveguides can also be employed in other types of microwave transmission lines. To illustrate this possibility, Section 8 describes Faraday rotators, directional phase shifters, circulators, and isolators in dielectric waveguides.

In Section 9 are indicated some of the advantages which may be obtained by employing isolators in waveguide circuits. The application of circulators to channel frequency filtering systems is described in connection with Figs. 69–71, and several forms of phase changer utilizing Faraday rotation or birefringence in round waveguide are described.

Not mentioned explicitly in earlier portions of this paper are the numerous possibilities for using ferrite-loaded circuits with time-varying magnetizing fields. The various directional or reciprocal phase changers, the isolators and the circulators which have been described may each be employed in conjunction with a modulating current which varies the magnetization on the ferrite and which thereby switches the transmission

properties of the ferrite-loaded circuit from one form to another. For example, an isolator may act as a matched-impedance On-Off switch simply by reversing the direction of the magnetizing field. Similarly, the direction of power circulation between the various terminals of a circulator may be reversed by reversing the direction of the magnetization. Continuous variation of the power division may in many of the circulators be achieved through variation of the magnetizing field.

It may be helpful to indicate the relative merits of several of the isolators which have previously been described. Both the Faraday rotation isolator and the resistance sheet isolator have produced comparable ratios of reverse-loss to forward-loss (about 150). The resonance isolator and the resistance-sheet isolator are simpler in form than Faraday rotation isolators and therefore may prove less expensive. The Faraday rotation isolator, and the resonance isolator, have provided reverse-loss to forward-loss ratios in the region 20–35 over frequency bands on the order of 10 per cent, whereas the resistance sheet isolator has provided ratios of 150 over the same band. However, the Faraday rotation isolator has not been perfected to the degree which appears possible with regard to broad-band operation. In comparing the resonance isolator with the resistance sheet isolator, the latter appears to have considerable advantage in requiring smaller magnetizing fields. This advantage of the resistance sheet isolator over the resonance isolator is more pronounced at the higher frequencies (such as 24,000 mc) where the magnet weight necessary to produce ferromagnetic resonance becomes objectionable.

REFERENCES

1. D. Polder, On the Theory of Electromagnetic Resonance, Phil. Mag., **40,** pp. 99–115, (1949).
2. H. G. Beljers, Physica, **16,** pp. 75–76, Jan., 1950.
3. F. F. Roberts; Jl. Phys. et Radium, **12,** p. 305, March, 1951.
4. For a list of references on ferrites and microwave Faraday rotation see: C. L. Hogan, The Ferromagnetic Faraday Effect at Microwave Frequencies and Its Applications, B.S.T.J., Jan., 1952; R. E. Alley, Jr., A Review of New Magnetic Phenomena, B.S.T.J., Sept., 1953.

5. B. D. H. Tellegen, Philips Res. Reports, **3**, pp. 81–101, (1948); **3**, pp. 321–332 (1948); **4**, pp. 31–37, (1949); **4**, pp. 366–369 (1949).
6. The "point-field" approach is due to A. G. Fox and was first published in Rev. Mod. Phys., **25**, pp. 262–263, Jan., 1953.
7. C. Kittel, Phys. Rev., **73,** p. 155, (1948).
8. R. L. Kyhl, M.I.T. Radiation Laboratory Series, **11,** Chapter 14, p. 887, also S. E. Miller, Coupled-Wave Theory and Waveguide Applications, B.S.T.J., **33,** pp. 661–719, May, 1954.
9. A similar argument is given by N. H. Frank, Introduction to Electricity and Optics, McGraw Hill, 1950, p. 344.
10. H. Suhl and L. R. Walker, Topics in Guided Wave Propagation Through Gyromagnetic Media — Part III, B.S.T.J., **33**, p. 1142, Equation 11, Sept. 1954.
11. Private communication. We are indebted to H. Seidel for permission to use his unpublished results and to S. P. Morgan for discussions regarding them.
12. G. C. Southworth, Principles and Applications of Waveguide Transmission, p. 130, Van Nostrand, 1950.
13. J. H. Rowen, Ferrites in Microwave Applications, B.S.T.J., **32**, p. 1358, Nov., 1953.
14. C. Kittel, Introduction to Solid State Physics, p. 164, J. Wiley, 1953.
15. D. W. Healy, Jr., Ferromagnetic Resonance in Nickel Ferrite, Phys. Rev., **86,** p. 1012, Fig. 6, June 15, 1952.
16. H. Suhl and L. R. Walker, Topics in Guided Wave Propagation through Gyromagnetic Media — Part I, B.S.T.J., **33,** Fig. 10(b), May, 1954.
17. C. L. Hogan, Reference 4, p. 25. A similar device was later described by C. H. Luhrs and Wm. J. Tull in U. S. Patent 2,644,930, issued July 7, 1953.
18. J. P. Schafer, Ferrite Isolators at 11,000 mc, scheduled to be published in Bell Laboratories Record.
19. The first published data on non-reciprocal phase shifts in rectangular waveguide was given by N. G. Sakiotis and H. N. Chait, Properties of Ferrites in Waveguides, Trans. of I.R.E., PGMTT, Vol. MTT-1, Nr. 2, pp. 11–16, Nov., 1953.
20. Principles of Microwave Circuits, Rad. Lab. Series, **8,** McGraw-Hill, pp. 385–388 (Note: equation (49), p. 387 is in error).
21. H. Suhl and L. R. Walker, Guided Wave Propagation Through Gyromagnetic Media — Part II, B.S.T.J., **33,** p. 954, July, 1954.
22. H. J. Riblet, The Short-Slot Hybrid Junction, Proc. I.R.E., **40,** pp. 180–184, Feb., 1952.
23. M. T. Weiss and A. G. Fox, Magnetic Double Refraction at Microwave Frequencies, Phys. Rev., **88,** pp. 146–147, Oct. 1, 1952. Note that equation (27) is written for γ as a negative quantity, although in the above reference γ was assumed positive.
24. E. H. Turner, A New Non-Reciprocal Waveguide Medium Using Ferrites, Proc. I.R.E., **41,** p. 937, July, 1953.
25. A. G. Fox, U. S. Patent, No. 2,531,419.
26. W. D. Lewis and L. C. Tillotson, A Non-Reflecting Branching Filter for Microwaves, B.S.T.J., **27,** p. 86, Jan., 1948.
27. A. G. Fox, An Adjustable Waveguide Phase Changer, Proc. I.R.E., **35,** pp. 1489–1498, Dec., 1947.
28. C. W. Tolbert and A. W. Straiton, Phase Shift Refractometer, Rev. Sci. Inst., **22,** p. 162, March, 1951.
29. D. H. Ring, A Microwave Double-Detection Measuring System with a Single Oscillator, presented at the High-Frequency Measurements Conference, Jan., 1953.

Correction

J. L. Merrill, Jr., A. F. Rose, and J. O. Smethurst, authors of the paper, "Negative Impedance Telephone Repeaters", which appeared in the September Issue of the B.S.T.J. on pages 1055 to 1092 have brought the following correction to the attention of the editors.

In Fig. 15, page 1074, the value of β_1 and β_2 were shown as follows:

$$\beta_1 = \frac{c}{a + b + c}$$

$$\beta_2 = \frac{a}{a + b + c}$$

This should be corrected to read:

$$\beta_1 = \frac{a}{a + b + c}$$

$$\beta_2 = \frac{c}{a + b + c}$$

The Potentials of Infinite Systems of Sources and Numerical Solutions of Problems in Semiconductor Engineering

By ARTHUR UHLIR, JR.

Tables and charts are given of mathematical functions related to the potential of a line of point charges. The use of these functions is illustrated by applications to semiconductor resistivity measurements and to calculations of the base resistance of point-contact transistors.

INTRODUCTION

The "method of images" is a simple but elegant technique for the solution of problems in potential theory. Formal solutions to problems involving plane boundaries and point sources can usually be written down immediately from obvious symmetry considerations. The solutions thus obtained, however, are often infinite series whose convergence is unsatisfactory for numerical calculations.

Many techniques have been devised for improving upon the convergence of these infinite sums. It has generally been assumed that the mathematician or engineer must select one of these techniques and apply it to his own special problem. The object of this article is to make such special treatment unnecessary in most cases. To this end, the potentials of certain simple image systems are tabulated.

The use of the tabulations is illustrated by numerical solutions of some problems in semiconductor engineering. Transistors and varistors have become famous because of their non-ohmic properties. But solutions to ohmic flow problems can be useful as first approximations in the design of semiconductor devices and can be very accurate approximations for suitably arranged measurements of the equilibrium resistivity of semiconductors.[1,2] Plane boundaries occur as the physical boundaries

of the semiconductor and as junctions between regions of different con-
ductivities and conductivity types. Point sources are the simplest
idealizations of point contacts.

The four-point method to be described for the measurement of semi-
conductor resistivity is used in geophysics for measuring the resistivity
of the earth. The mathematical content of this article is applicable to
problems other than ohmic flow. The differential equations for ohmic
flow are identical with those of electrostatics, heat conduction, and the
hydrodynamics of an ideal fluid. In one of the examples given below,
the Coulomb energy of an ionic crystal is calculated.

SUPERPOSITION OF AN INFINITE NUMBER OF POINT SOURCES

The strength q of a point source will be defined so that the potential
due to the source is q/r, where r is the distance from the source. Sources
of negative strength are often called "sinks", but this special designation
will not be used in the discussion to follow. All the cases of ohmic flow to
be considered here involve point sources of current at a plane surface
and it will be shown that

$$q = \frac{I\rho}{2\pi} \qquad (1)$$

where I is the current into the plane surface of material of resistivity ρ.

According to the principle of superposition, the potential at a given
point due to some configuration of fixed sources is the sum of the po-
tentials that would be established at that point by the various sources
making up the configuration. The computations for many problems
could be accomplished easily if one could have a table of the potential
of a line of equally-spaced point sources of equal value q. But a divergent
expression is obtained when one attempts to calculate the potential of
such an arrangement of sources by superposing the q/r potentials of
the sources. One way to avoid this divergence is to abandon the attempt
(implicit in the choice of q/r as the potential of a source) to make the
potential at infinity equal to zero. Potentials with respect to some ar-
bitrary point would be finite and could be tabulated.

However, having the potential at infinity equal to zero simplifies the
superposition of source systems. All of the problems that could be solved
with the potential of a line of point sources can be solved with either of
two related arrangements that permit setting the potential at infinity
equal to zero. Tabulations of the potentials of these arrangements will
be given.

The first scheme that will be considered consists entirely of point sources and is illustrated in Fig. 1. It consists of a line of equally spaced point sources of equal strength q, and a parallel line of opposite sources, except that one of the positive sources and the corresponding negative source have been omitted. The potential at the point P (where the omitted positive source would have been placed) is to be considered and may be written

$$V(P) = \frac{q}{a} M(\lambda) \tag{1}$$

where M is a dimensionless function of the ratio $\lambda = d/a$ of the distance d between the lines of sources to the spacing a of the charges in the lines. It is obvious that $M(\lambda)$ must vanish at $\lambda = 0$.

The second scheme is illustrated in Fig. 2. It consists of a line of equally spaced point sources embedded in a line source of equal and opposite strength per unit length. The potential at the point Q, lying on a perpendicular to the line of charges, erected at one of the point sources
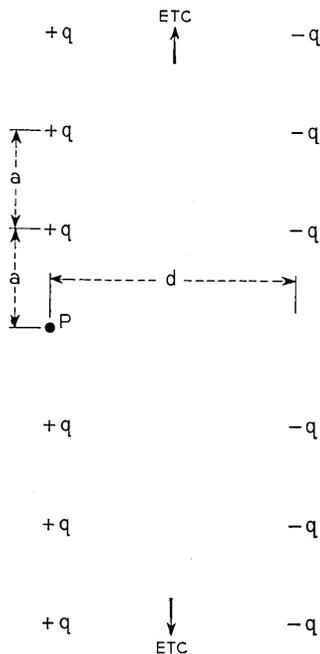
Fig. 1 — Arrangement of sources corresponding to the function M. The potential at point P is $(q/a) M(d/a)$.
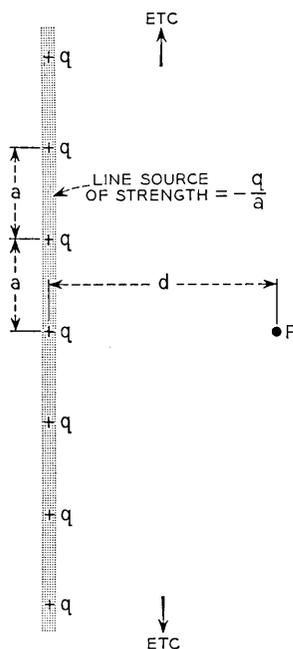
Fig. 2 — Source distribution corresponding to the function N. The potential at point P is $(q/a) N(d/a)$.
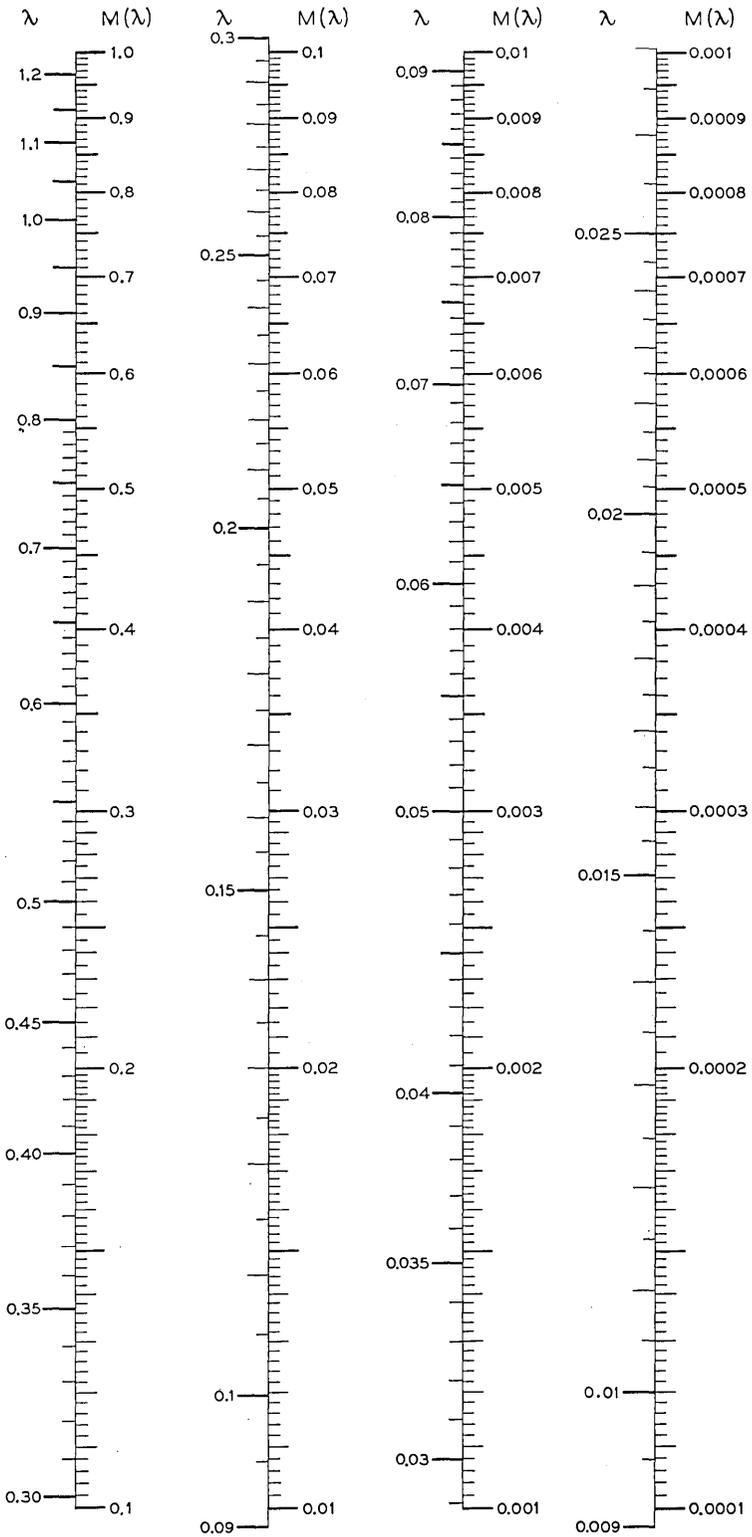
Fig. 3 — The function $M(\lambda) = 2 \sum\limits_{n=1}^{\infty} \left(\dfrac{1}{n} - \dfrac{1}{\sqrt{n^2 + \lambda^2}} \right)$.
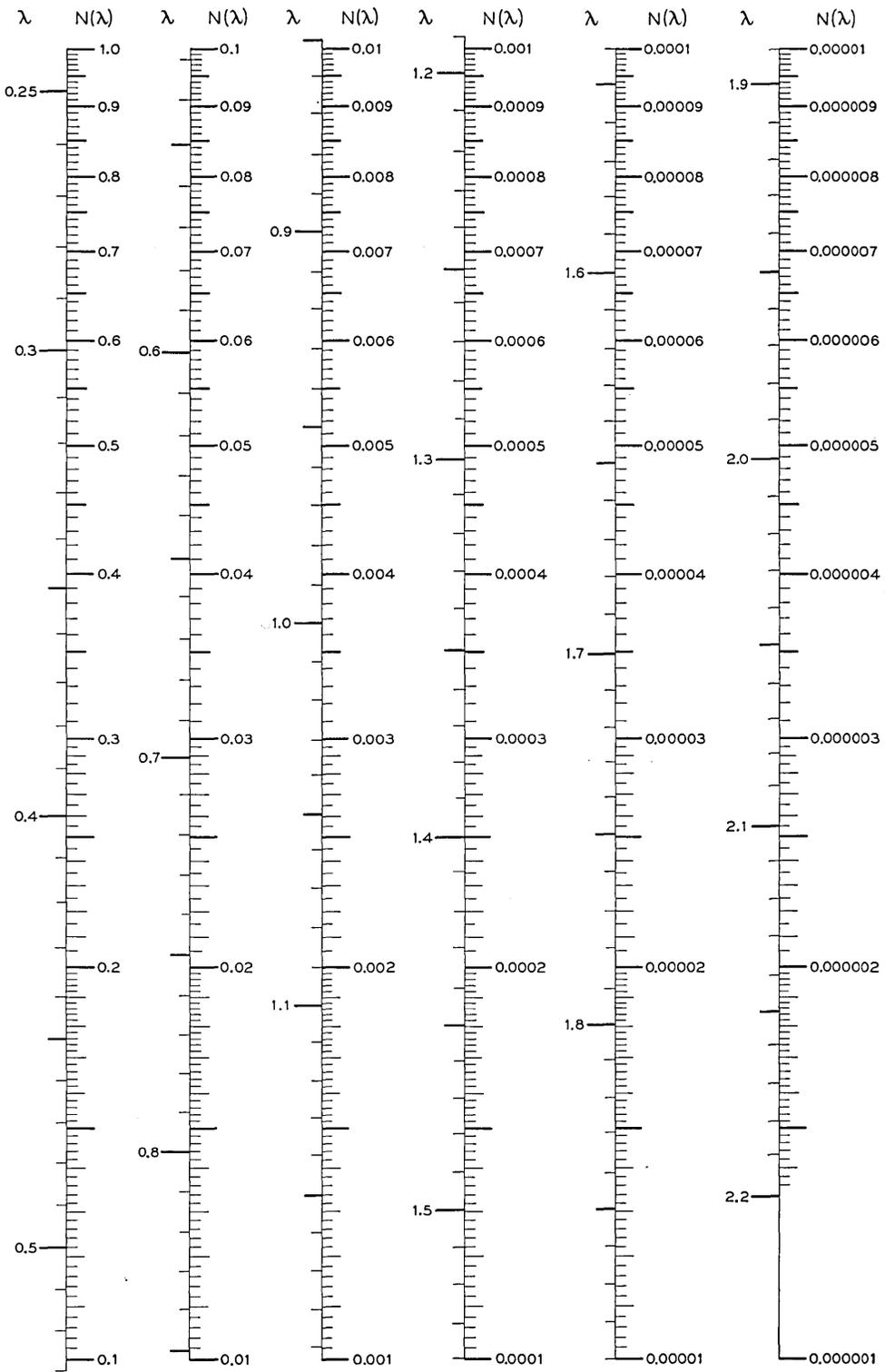
Fig. 4 — The function $N(\lambda) = 2\pi \sum_{n=1}^{\infty} iH_0^{(1)}(i2\pi n\lambda)$.

Table I — Values of the Functions $M(\lambda)$ and $N(\lambda)$

λ	$M(\lambda)$	$N(\lambda)$	λ	$M(\lambda)$	$N(\lambda)$
0.01	0.000120	90.557677	0.52	0.278275	0.105086
0.02	0.000481	41.943612	0.53	0.287536	0.097637
0.03	0.001081	26.087274	0.54	0.296876	0.090740
0.04	0.001921	18.328465	0.55	0.306292	0.084353
0.05	0.003000	13.773673	0.56	0.315780	0.078434
0.06	0.004317	10.803665	0.57	0.325338	0.072947
0.07	0.005872	8.729459	0.58	0.334961	0.067860
0.08	0.007662	7.209018	0.59	0.344647	0.063140
0.09	0.009686	6.053671	0.60	0.354392	0.058761
0.10	0.011943	5.151023			
			0.61	0.364193	0.054696
0.11	0.014432	4.430064	0.62	0.374047	0.050921
0.12	0.017150	3.843793	0.63	0.383952	0.047416
0.13	0.020096	3.359908	0.64	0.393903	0.044159
0.14	0.023266	2.955502	0.65	0.403900	0.041133
0.15	0.026660	2.613904	0.66	0.413937	0.038320
0.16	0.030273	2.322701	0.67	0.424014	0.035705
0.17	0.034105	2.072472	0.68	0.434127	0.033273
0.18	0.038151	1.855945	0.69	0.444273	0.031012
0.19	0.042410	1.667423	0.70	0.454451	0.028907
0.20	0.046877	1.502384			
			0.71	0.464658	0.026949
0.21	0.051550	1.357196	0.72	0.474890	0.025127
0.22	0.056426	1.228910	0.73	0.485148	0.023431
0.23	0.061501	1.115110	0.74	0.495426	0.021852
0.24	0.066773	1.013798	0.75	0.505725	0.020381
0.25	0.072237	0.923312	0.76	0.516041	0.019012
0.26	0.077889	0.842254	0.77	0.526373	0.017736
0.27	0.083726	0.769448	0.78	0.536718	0.016548
0.28	0.089746	0.703889	0.79	0.547074	0.015440
0.29	0.095942	0.644722	0.80	0.557441	0.014409
0.30	0.102312	0.591213			
			0.81	0.567816	0.013447
0.31	0.108852	0.542726	0.82	0.578196	0.012551
0.32	0.115557	0.498711	0.83	0.588582	0.011715
0.33	0.122425	0.458690	0.84	0.598970	0.010936
0.34	0.129450	0.422244	0.85	0.609360	0.010210
0.35	0.136629	0.389006	0.86	0.619750	0.009532
0.36	0.143958	0.358654	0.87	0.630138	0.008900
0.37	0.151432	0.330903	0.88	0.640523	0.008311
0.38	0.159048	0.305500	0.89	0.650904	0.007761
0.39	0.166801	0.282222	0.90	0.661279	0.007248
0.40	0.174687	0.260868			
			0.91	0.671647	0.006770
0.41	0.182703	0.241262	0.92	0.682007	0.006323
0.42	0.190844	0.223244	0.93	0.692358	0.005906
0.43	0.199107	0.206671	0.94	0.702698	0.005518
0.44	0.207486	0.191417	0.95	0.713027	0.005155
0.45	0.215979	0.177364	0.96	0.723344	0.004816
0.46	0.224582	0.164411	0.97	0.733647	0.004499
0.47	0.233289	0.152462	0.98	0.743935	0.004204
0.48	0.242098	0.141434	0.99	0.754209	0.003928
0.49	0.251005	0.131248	1.00	0.764466	0.003671
0.50	0.260006	0.121836			
			1.01	0.774706	0.003431
0.51	0.269098	0.113135	1.02	0.784928	0.003206

TABLE I — VALUES OF THE FUNCTIONS $M(\lambda)$ AND $N(\lambda)$ — (*Continued*)

λ	$M(\lambda)$	$N(\lambda)$	λ	$M(\lambda)$	$N(\lambda)$
1.03	0.795132	0.002996	1.54	1.280952	0.000100
1.04	0.805316	0.002801	1.55	1.289715	0.000093
1.05	0.815480	0.002618	1.56	1.298447	0.000087
1.06	0.825624	0.002447	1.57	1.307149	0.000082
1.07	0.835746	0.002287	1.58	1.315821	0.000077
1.08	0.845847	0.002138	1.59	1.324464	0.000072
1.09	0.855924	0.001999	1.60	1.333077	0.000067
1.10	0.865979	0.001869			
			1.61	1.341660	0.000063
1.11	0.876010	0.001748	1.62	1.350214	0.000059
1.12	0.886018	0.001634	1.63	1.358739	0.000055
1.13	0.896000	0.001528	1.64	1.367234	0.000051
1.14	0.905958	0.001428	1.65	1.375700	0.000048
1.15	0.915890	0.001336	1.66	1.384137	0.000045
1.16	0.925797	0.001249	1.67	1.392544	0.000042
1.17	0.935677	0.001168	1.68	1.400923	0.000040
1.18	0.945531	0.001092	1.69	1.409273	0.000037
1.19	0.955358	0.001022	1.70	1.417594	0.000034
1.20	0.965158	0.000956			
			1.71	1.425886	0.000032
1.21	0.974930	0.000894	1.72	1.434150	0.000030
1.22	0.984675	0.000836	1.73	1.442385	0.000028
1.23	0.994392	0.000782	1.74	1.450593	0.000026
1.24	1.004080	0.000731	1.75	1.458772	0.000025
1.25	1.013740	0.000684	1.76	1.466923	0.000024
1.26	1.023371	0.000640	1.77	1.475046	0.000022
1.27	1.032974	0.000599	1.78	1.483141	0.000020
1.28	1.042547	0.000560	1.79	1.491208	0.000019
1.29	1.052091	0.000524	1.80	1.499248	0.000018
1.30	1.061606	0.000490			
			1.81	1.507260	0.000017
1.31	1.071091	0.000459	1.82	1.515244	0.000016
1.32	1.080547	0.000429	1.83	1.523202	0.000015
1.33	1.089973	0.000401	1.84	1.531132	0.000014
1.34	1.099369	0.000376	1.85	1.539036	0.000013
1.35	1.108735	0.000351	1.86	1.546912	0.000012
1.36	1.118072	0.000329	1.87	1.554762	0.000011
1.37	1.127378	0.000308	1.88	1.562585	0.000010
1.38	1.136654	0.000288	1.89	1.570381	0.000010
1.39	1.145900	0.000270	1.90	1.578151	0.000009
1.40	1.155115	0.000252			
			1.91	1.585895	0.000009
1.41	1.164300	0.000236	1.92	1.593612	0.000008
1.42	1.173455	0.000221	1.93	1.601304	0.000008
1.43	1.182580	0.000207	1.94	1.608970	0.000007
1.44	1.191674	0.000193	1.95	1.616609	0.000007
1.45	1.200738	0.000181	1.96	1.624224	0.000006
1.46	1.209772	0.000169	1.97	1.631812	0.000006
1.47	1.218775	0.000159	1.98	1.639376	0.000006
1.48	1.227749	0.000148	1.99	1.646913	0.000005
1.49	1.236691	0.000139	2.00	1.654426	0.000005
1.50	1.245604	0.000130			
			2.10	1.728200	0.000003
1.51	1.254486	0.000122	2.20	1.799596	0.000001
1.52	1.263338	0.000114	2.30	1.868737	0.000001
1.53	1.272161	0.000106	2.40	1.935741	0.000000
			2.50	2 000718	0.000000

may be written

$$V(Q) = \frac{q}{a} N(\lambda) \tag{2}$$

if we continue to denote by a the spacing between the point sources and now take d to be the distance between the point Q and the line. As λ becomes infinite, N goes to zero.

The potential of a line of point sources along a perpendicular to the line erected at one of the point sources can be taken at

$$V = \frac{q}{a} \left[\frac{1}{\lambda} - M(\lambda) \right] \tag{3}$$

where $\lambda = d/a$ and d is the distance from the line of sources.

Numerical values of M and N can be obtained with sufficient accuracy for most purposes from the charts in Figs. 3 and 4. More accurate values of M and N are given in Table I for various values of λ.

A useful relation exists between the functions M and N. It is

$$M(\lambda) + N(\lambda) = \frac{1}{\lambda} + 2 \ln \lambda - 2 \ln 2 + 2C, \tag{4}$$

where C is Euler's constant, 0.577215665. That is

$$M(\lambda) + N(\lambda) = \frac{1}{\lambda} + 2 \ln \lambda - 0.231863031$$

The methods of computing M and N are described in the appendix.
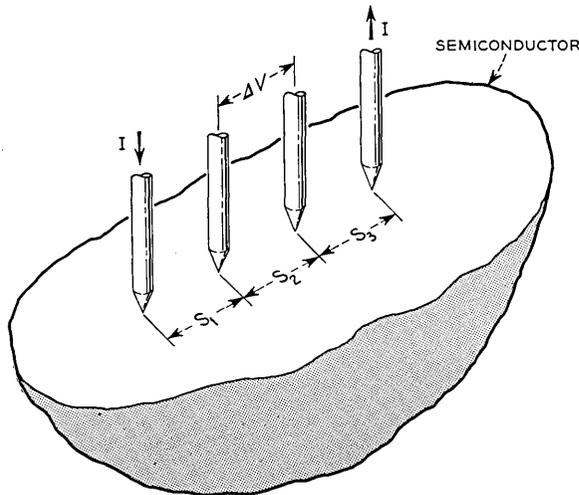
Fig. 5 — Linear four-point probe.

FOUR-POINT RESISTIVITY MEASUREMENTS

Resistivity measurements are regularly being made on germanium with a four-point probe. The basic assumptions are that the resistivity is uniform and that the potential distribution can be accurately approximated by that calculated for ohmic flow. The results of calculations for some geometries have been presented by L. B. Valdes, along with a discussion of the conditions for the validity of the assumption of ohmic flow.[1] In brief, it is required that modulation of the conductivity by injected minority carriers be kept negligible. It is also necessary that the probes used to measure potential difference be sufficiently conducting to pass the current drawn by the voltage measuring instrument. Both of these requirements can usually be met, in the case of germanium, by abrading the surface or by subjecting the contact between probe and germanium to an electric discharge, *i.e.*, "forming" the contact.

The prototype geometry for four-point resistivity measurements is given in Fig. 5, which is taken from Reference 1. The potentials at probes 2 and 3 are

$$V_2 = \frac{q}{S_1} - \frac{q}{S_2 + S_3}$$

$$V_3 = \frac{q}{S_1 + S_2} - \frac{q}{S_3}$$

(5)

where q is the strength of a source corresponding to the current I. By By considering a hemispherical surface of infinitesimal radius r centered on one of the current probes, one sees that

$$I = \frac{1}{\rho} \times \text{area} \times E = \frac{1}{\rho} 2\pi r^2 \frac{q}{r^2}$$

or

$$q = I\rho/2\pi \tag{6}$$

where ρ is the resistivity and $E = q/r^2$ is the magnitude of the electric field.

The potential difference ΔV between probes 2 and 3 is

$$\Delta V = q\left(\frac{1}{S_1} + \frac{1}{S_3} - \frac{1}{S_2 + S_3} - \frac{1}{S_1 + S_2}\right) \tag{7}$$

It will be assumed in the following examples that $S_1 = S_2 = S_3 = s$ in which case

$$\Delta V = q/s \tag{8}$$

so that

$$\rho = 2\pi s(\Delta V/I). \tag{9}$$

When the four-point probe is applied to a solid that is not approximately semi-infinite, a potential difference will still be observed when a current flows, so we will define an "apparent resistivity" ρ_0 by

$$\rho_0 = 2\pi s(\Delta V/I) \tag{10}$$

The true resistivity will be given by

$$\rho = \rho_0/\text{C.D.} \tag{11}$$

where C.D. is the correction divisor for the particular case.

One of the simplest and most useful departures from the semi-infinite geometry is the slab or slice of finite thickness, free of any conducting coating on its faces (and resting on a non-conducting support during the measurement). Fig. 6 shows the position of the probes and the configuration of sources required for calculating the correction divisor. The required configuration can be obtained by superimposing two of the arrangements shown in Fig. 1 (one with opposite sign) i.e., upon the source and sink used for the semi-infinite geometry. The additional potential at one of the probes due to the additional sources is

$$\frac{q}{2W}\left[M(2s/2w) - M(s/2w)\right]$$

and the additional potential at the other probe is equal and opposite, so that the additional potential difference is

$$\frac{q}{w}\left[M(2s/2w) - M(s/2w)\right]$$

In comparison to the potential difference for the semi-infinite case, we find that ΔV is "too high" by a factor

$$\begin{aligned}
\text{C.D.} &= \frac{s}{q}\left(\frac{q}{s} + \frac{q}{w}\left[M\left(\frac{2s}{2w}\right) - M\left(\frac{s}{2w}\right)\right]\right) \\
&= 1 + \frac{s}{w}\left[M\left(\frac{s}{w}\right) - M\left(\frac{s}{2w}\right)\right]
\end{aligned} \tag{12}$$

This expression is true in general, but is most convenient for slices that are thick in comparison to the spacing of the probes, since the correction terms are then small and errors in reading M from Fig. 3 do not matter much. The correction divisor may be given in terms of the function N

through use of equation (4). The result,

$$\text{C.D.} = \frac{s}{w}\left[2\ln 2 + N\left(\frac{s}{2w}\right) - N\left(\frac{s}{w}\right)\right] \tag{13}$$

is most useful for thin slices. In fact, if the slice thickness w is less than half the probe spacing s, C.D. $\approx (2\ln 2)\ s/w$ and equations (10) and (11) give

$$\rho \approx \pi w \Delta V/I \ln 2 \approx 4.53 w \Delta V/I \qquad \text{(for } w \ll \frac{1}{2}s) \tag{14}$$

The correction divisor is given in a table in Fig. 6 for a number of ratios of probe spacing to slice thickness. For larger values of this ratio than are

s/w	C.D.	s/w	C.D.	s/w	C.D.
0.1	1.0009	0.6	1.1512	1.2	1.7329
0.2	1.0070	0.7	1.2225	1.4	1.9809
0.3	1.0227	0.8	1.3062	1.6	2.2410
0.4	1.0511	0.9	1.4008	1.8	2.5083
0.5	1.0939	1.0	1.5045	2.0	2.7799
				2.5	3.4674
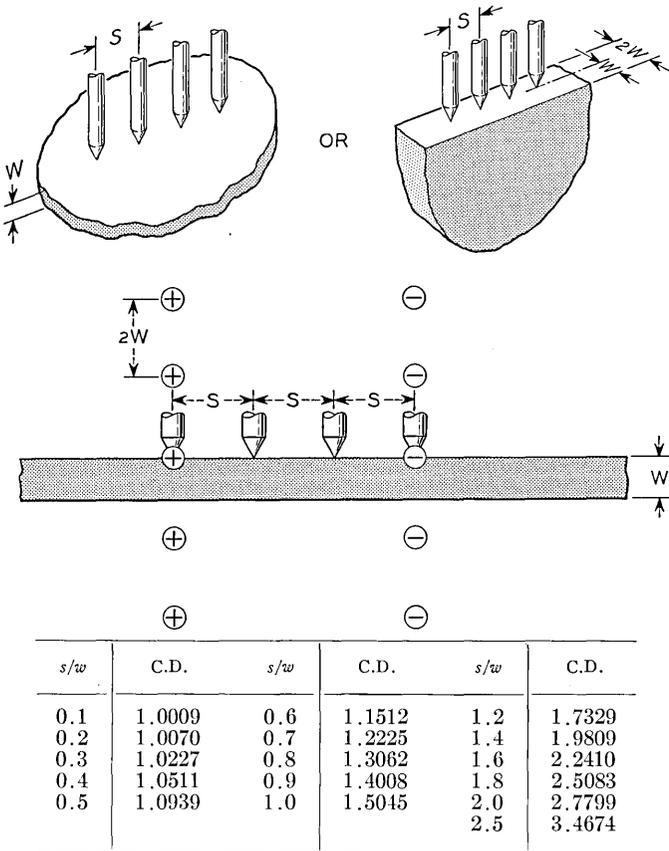
Fig. 6 — Linear probe on infinite slice with nonconducting faces (or on edge of semi-infinite slice).

found in the table, equation (14) may be used with sufficient accuracy for all practical resistivity measurements.

In general, a large correction divisor is desirable because it indicates that a relatively large potential difference is to be measured. For this reason, the use of the four-point probe on thin slices should be and, in fact, is found in practice to be quite satisfactory. However, if the slice is provided with a conducting coating on one side, as in intermediate stages of point-contact transistor fabrication, the measurement is apt to be inaccurate. The correction divisor for this case is

$$\text{C.D.} = 1 + \frac{s}{w}\left[2M\left(\frac{s}{2w}\right) - M\left(\frac{s}{w}\right) - M\left(\frac{s}{4w}\right)\right] \qquad (15)$$

or

$$\text{C.D.} = \frac{s}{w}\left[N\left(\frac{s}{4w}\right) + N\left(\frac{s}{w}\right) - 2N\left(\frac{s}{2w}\right)\right] \qquad (16)$$

Fig. 7 illustrates this configuration and contains a table of the correction divisor. It will be noted that the correction divisor is quite small if the
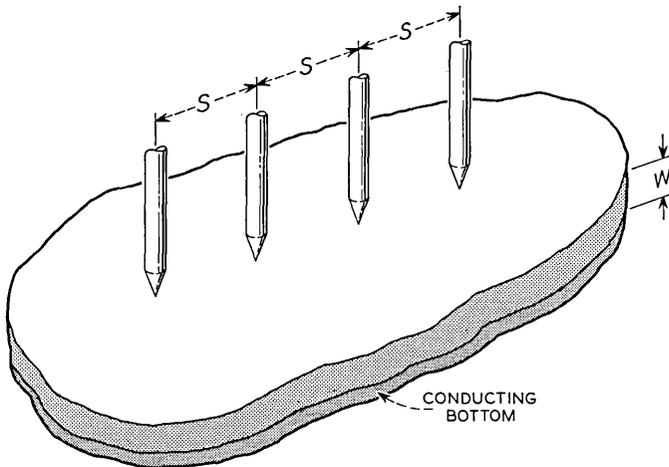
s/w	C.D.	s/w	C.D.	s/w	C.D.	s/w	C.D.
0.1	.9993	0.5	.9329	1.0	.6833	2.0	.2283
0.2	.9948	0.8	.7960	1.5	.4159	5.0	.0034

Fig. 7 — Linear probe on infinite slice with conducting bottom face.
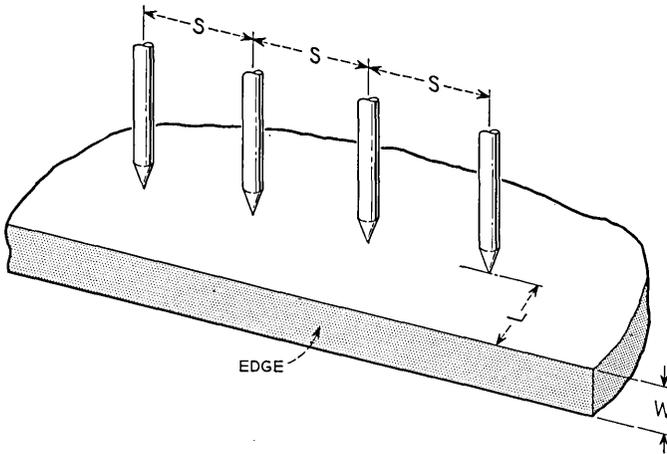
Correction Divisor for Conducting Case

L/s

	0.1	0.2	0.5	1.0	2.0	5.0	10.0
0.0	0.034	0.124	0.481	0.811	0.962	0.9971	0.9996
0.1	0.03	0.124	0.48	0.81	0.96	0.997	1.0001
0.2	0.03	0.125	0.48	0.81	0.96	1.002	1.0064
0.5	0.04	0.125	0.49	0.83	1.01	1.08	1.090
1.0	0.04	0.142	0.56	1.03	1.34	1.48	1.497
2.0	0.066	0.22	0.95	1.84	2.46	2.72	2.765
5.0	0.146	0.55	2.35	4.58	6.12	6.78	6.894

(leftmost column header: $\dfrac{s}{w}$)

Correction Divisor for Non Conducting Case

L/s

	0	0.1	0.2	0.5	1.0	2.0	5.0	10.0
0.0	2.000	1.9661	1.8764	1.5198	1.1890	1.0379	1.0029	1.0004
0.1	2.002	1.97	1.88	1.52	1.19	1.040	1.004	1.0017
0.2	2.016	1.98	1.89	1.53	1.20	1.052	1.014	1.0094
0.5	2.188	2.15	2.06	1.70	1.35	1.176	1.109	1.0977
1.0	3.009	2.97	2.87	2.45	1.98	1.667	1.534	1.512
2.0	5.560	5.49	5.34	4.61	3.72	3.104	2.838	2.795
5.0	13.863	13.72	13.32	11.51	9.28	7.744	7.078	6.969
10.0	27.726	27.43	26.71	23.03	18.56	15.49	14.156	13.938

(leftmost column header: $\dfrac{s}{w}$)

Fig. 8 — Semi-infinite slice — linear probe parallel to edge.

spacing between probe points is much larger than the slice thickness. The two-point measurement to be described below seems preferable to the four-point measurement in the case of thin slices with a conducting coating, because less effort is required to maintain an accurate small spacing between two points than between four points.

The two cases that have been treated thus far were also given by Valdes, but have been mentioned here as simple examples of the use of the M and N functions. An example of a more complicated problem is that of the linear four-point probe on a semi-infinite slice, parallel to the edge of the slice, as shown in Fig. 8. It is useful to note that the
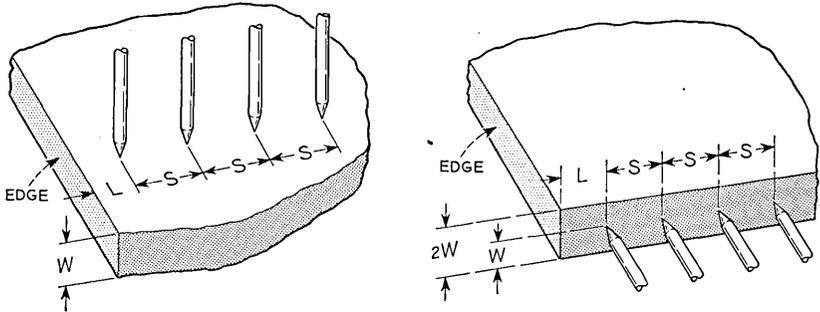
Correction Divisor for Nonconducting Edge

s/w	L/s								
	0	0.1	0.2	0.5	1.0	2.0	5.0	10.0	∞
0.0	1.4500	1.3330	1.2555	1.1333	1.0595	1.0194	1.0028	1.0005	1.0000
0.1	1.4501	1.3331	1.2556	1.1335	1.0597	1.0198	1.0035	1.0015	1.0009
0.2	1.4519	1.3352	1.2579	1.1364	1.0637	1.0255	1.0107	1.0084	1.0070
0.5	1.5285	1.4163	1.3476	1.2307	1.1648	1.1263	1.1029	1.0967	1.0939
1.0	2.0335	1.9255	1.8526	1.7294	1.6380	1.5690	1.5225	1.5102	1.5045
2.0	3.7236	3.5660	3.4486	3.2262	3.0470	2.9090	2.8160	2.7913	2.7799
5.0	9.2815	8.8943	8.6025	8.0472	7.5991	7.2542	7.0216	6.9600	6.9315
10.0	18.5630	17.7886	17.2050	16.0944	15.1983	14.5083	14.0431	13.9199	13.8629

Correction Divisor for Conducting Edge

s/w	L/s								
	0	0.1	0.2	0.5	1.0	2.0	5.0	10.0	∞
0.0	0.5500	0.6670	0.7445	0.8667	0.9405	0.9806	0.9972	0.9995	1.0000
0.1	0.5517	0.6687	0.7462	0.8683	0.9421	0.9820	0.9982	1.0003	1.0009
0.2	0.5620	0.6788	0.7560	0.8775	0.9502	0.9885	1.0033	1.0056	1.0070
0.5	0.6593	0.7714	0.8402	0.9571	1.0230	1.0615	1.0849	1.0910	1.0939
1.0	0.9754	1.0835	1.1563	1.2796	1.3709	1.4399	1.4864	1.4988	1.5045
2.0	1.8362	1.9938	2.1113	2.3336	2.5129	2.6508	2.7439	2.7685	2.7799
5.0	4.5815	4.9687	5.2605	5.8158	6.2638	6.6088	6.8413	6.9030	6.9315
10.0	9.1629	9.9373	10.5209	11.6315	12.5276	13.2176	13.8060	13.8060	13.8629

Fig. 9 — Semi-infinite slice with probe perpendicular to edge (or quarter-infinite slice with probe on a nonconducting edge).

results from Fig. 6 can be incorporated into the calculations, for we may let the correction divisor

$$\text{C.D.} = \text{C.D. (for nonconducting infinite slices)} \pm \Delta\text{C.D.}, \quad (17)$$

where the plus sign is used if the edge is nonconducting and the minus sign if the edge is conducting (the faces of the slice are assumed to be nonconducting in either case). Then

$$\Delta\text{C.D.} = \frac{1}{\sqrt{\eta^2 + \frac{1}{4}}} - \frac{1}{\sqrt{\eta^2 + 1}} \\ + \mu[M(\mu\sqrt{\eta^2 + 1}) - M(\mu\sqrt{\eta^2 + \frac{1}{4}})] \quad (18)$$

or

$$\Delta\text{C.D.} = \mu\left[\ln\frac{\eta^2 + 1}{\eta^2 + \frac{1}{4}} + N(\mu\sqrt{\eta^2 + \frac{1}{4}}) - N(\mu\sqrt{\eta^2 + 1})\right] \quad (19)$$

where

$$\mu = \frac{s}{w} \quad \text{and} \quad \eta = \frac{L}{s} \quad (20)$$

As shown in Fig. 8, L is the distance from the edge to the probe, s is the spacing of the probes, and w is the slice thickness. A table of values of the correction divisor is given in Fig. 8.

The situation shown in Fig. 9 is similar to that of Fig. 8 in that the linear four-point probe is near the edge of a semi-infinite slice. But in this case the probe is perpendicular to the edge and L is the distance from the edge to the nearest point of the probe. Here again the correction divisor can be written in terms of the result for infinite slices, equation (17). It is found that

$$\Delta\text{C.D.} = \frac{s}{2w}\left(\frac{1}{\alpha} + \frac{1}{\beta} - \frac{1}{\gamma} - \frac{1}{\delta} + M(\gamma) + M(\delta) - M(\alpha) - M(\beta)\right) \quad (21)$$

or

$$\Delta\text{C.D.} = \frac{s}{2w}\left(2\ln\frac{\gamma\delta}{\alpha\beta} + N(\alpha) + N(\beta) - N(\gamma) - N(\delta)\right) \quad (22)$$

where

$$\alpha = (L + \tfrac{1}{2}s)/w$$
$$\beta = (L + \tfrac{3}{2}s)/w$$
$$\gamma = (L + s)/w \qquad (23)$$
$$\delta = (L + 2s)/w$$

Correction divisors can be compounded still further. For example, the correction divisor for the quarter-infinite slice with probe on the diagonal, shown in Fig. 10, may be written to make use of the results for Fig. 9. That is,

$$\text{C.D.} = \text{C.D. (Semi-infinite slice with probe perpendicular to} \quad (24)$$
$$\text{non-conducting edge, see Fig. 9)} \pm \Delta\text{C.D.}$$

Again, the plus sign is taken if both edges are nonconducting, the minus sign if both edges are conducting. If α, β, γ, and δ are redefined by

$$\alpha = \frac{1}{2w} \sqrt{L^2 + (S + L)^2}$$

$$\beta = \frac{1}{2w} \sqrt{(L + 2s)^2 + (L + 3s)^2}$$

$$\gamma = \frac{1}{2w} \sqrt{L^2 + (L + 2s)^2} \tag{25}$$

$$\delta = \frac{1}{2w} \sqrt{(L + 3s)^2 + (L + S)^2}$$

then ΔC.D. can be calculated by inserting equation (25) into equation (21) or (22) and doubling the result. One numerical example will be
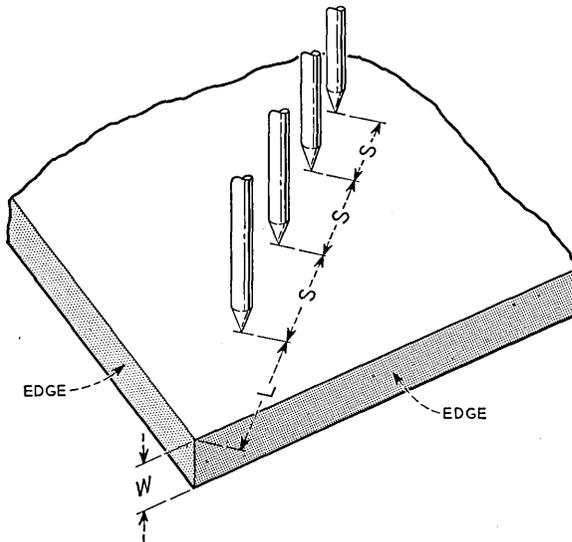
Fig. 10 — Linear probe on quarter-infinite slice.

given. If $s = L = w$, C.D. $= 2.1097$ for nonconducting edges and 1.1665 for conducting edges.

It is evident that tabulation of correction divisors for situations with two or more degrees of freedom is rather tedious. Yet the computation of a correction divisor for a given set of dimensions may be very simple, requiring nothing more than the reading of several values of M or N from Fig. 3 or Fig. 4.

In a four-point resistivity measurement it is not necessary that the points be equally spaced nor that they be colinear. Whatever the chosen arrangement, the correction divisor can be calculated from the M and N functions for any position of the probe on infinite, semi-infinite, or quarter-infinite slices. Numerous less frequently encountered situations may also be treated.

The square arrangement of probe points illustrated in Fig. 11 is useful. When used on the surface of a semi-infinite solid, the resistivity is given by

$$\rho = \rho_0 = \frac{2\pi b}{2 - \sqrt{2}} \frac{\Delta V}{I} \tag{26}$$
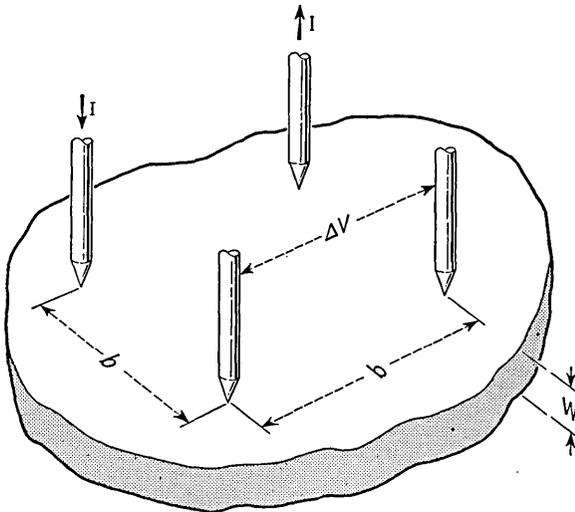
b/w	0.1	0.2	0.5	1.0	2.0	5.0	10.0
C.D.	1.0005	1.004	1.057	1.344	2.378	5.916	11.832

Fig. 11 — Square arrangement of points — infinite slice.

where b is the length of the side of the square. When the square probe is used on infinite slices with nonconducting faces, the resistivity calculated from equation (26) must be corrected by dividing by

$$\text{C.D.} = 1 + \frac{b}{(2 - \sqrt{2})w} \left[M\left(\frac{b}{\sqrt{2}w}\right) - M\left(\frac{b}{2w}\right) \right] \qquad (27)$$

or

$$\text{C.D.} = \frac{b}{(2 - \sqrt{2})w} \left[\ln 2 + N\left(\frac{b}{2w}\right) - N\left(\frac{b}{\sqrt{2}w}\right) \right] \qquad (28)$$

Some values of this correction divisor are given in Fig. 11. For thin slices,

$$\rho \approx \frac{2\pi w \Delta V}{I \ln 2} = 9.06w \frac{\Delta V}{I} \qquad \left(w \ll \frac{1}{2}b \right) \qquad (29)$$

BASE RESISTANCE OF A POINT CONTACT TRANSISTOR AND TWO-POINT RESISTIVITY MEASUREMENTS

The base resistance r_b of a transistor is defined by

$$r_b = (\partial V_\epsilon / \partial I_c)_{I\epsilon} \qquad (30)$$

where V_ϵ is the emitter voltage, I_c is the collector current, and I_ϵ is the emitter current. The main contribution to r_b for a wide-spaced point-contact transistor seems to be the ohmic resistivity of the semi-conductor. In close-spaced point-contact transistors, the widening of
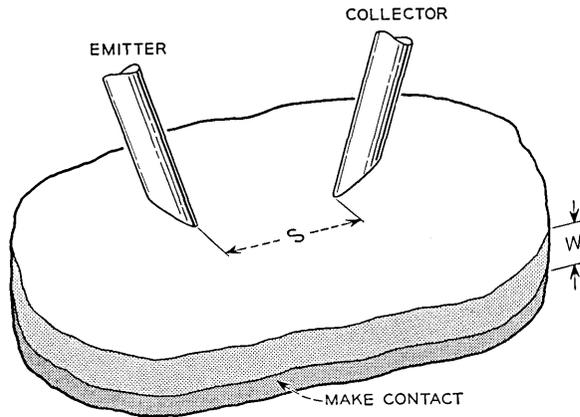
s/w	0.1	0.2	0.5	1.0	2.0
C.F.	0.931	0.862	0.667	0.401	0.118

Fig. 12 — Base resistance of a point-contact transistor.

the collector space-charge region and the modulation of conductivity by carrier injection have important effects on r_b; it is necessary to calculate the ohmic base resistance to evaluate these effects. The ohmic base resistance for a transistor made on an infinite slice with a base contact on the underside, as shown in Fig. 12, can be calculated from the M or N functions. It is found that

$$r_b = \frac{\rho}{2\pi s}\left(1 - \frac{s}{w}\ln 2 + \frac{s}{2w}\left[M\left(\frac{s}{2w}\right) - M\left(\frac{s}{4w}\right)\right]\right) \qquad (30)$$

or

$$r_b = \frac{\rho}{4\pi w}\left[N\left(\frac{s}{4w}\right) - N\left(\frac{s}{2w}\right)\right] \qquad (31)$$

These results can be expressed in terms of a correction factor:

$$r_b = \frac{\rho}{2\pi s} \times \text{C.F.} \qquad (32)$$

where $\rho/2\pi s$ is the ohmic base resistance for a transistor constructed on a semi-infinite solid with base contact at infinity. Some values of the correction factor are given in Fig. 12.

   In measuring the resistivity of a slice which has been provided with a conducting coating on one face, it is better to use two points in the geometry shown in Fig. 12 than to use a four-point probe, for the mechanical reason mentioned previously. The two-point probe also gives better localization of the measurement and reversal of the current affords a valuable check on the assumption of ohmic flow. Obviously, the correction *divisors* for correcting such a resistivity measurement for slice thickness are equal to the base resistance correction *factors* given in Fig. 12.

FURTHER APPLICATIONS OF THE TABULATED FUNCTIONS

   In the previous examples, the potential was required only along a line perpendicular to a line of point sources *and passing through one of the sources*. The tabulations are of more general usefulness. For example, $[M(\lambda) - 2M(2\lambda)]q/a$ gives the potential of a line of point sources along a perpendicular bisector of the line between adjacent sources (this potential is zero when $\lambda = 0$; the potential at infinity is $-\infty$.).

   The potential of a plane grid of equal point sources is easily obtained from the tabulations. Suppose that the sources have a regular spacing a in one direction and b in the perpendicular direction. There is no loss of generality in assuming that b is greater than or equal to a. The potential at a point a distance d from the plane of sources and on a perpendicular

to the plane erected at one of the sources is

$$V = \frac{q}{a} R_k(\lambda) \tag{34}$$

where $\lambda = d/a$, $k = b/a$, and

$$R_k(\lambda) = \frac{1}{\lambda} - M(\lambda) - \frac{1}{k} M\left(\frac{\lambda}{k}\right)$$

$$- 2 \sum_{n=1}^{\infty} \left[M\left(\sqrt{\lambda^2 + n^2 k^2}\right) - M(nk) \right] \tag{35}$$

Equation (4) enables one to transform equation (35) into the following rapidly converging expressions:

$$R_k(\lambda) = \frac{1}{\lambda} - M(\lambda) - 2 \ln \frac{\sinh \pi\lambda/k}{\pi\lambda/k}$$

$$- 2 \sum_{n=1}^{\infty} \left[N(nk) - N\left(\sqrt{\lambda^2 + n^2 k^2}\right) \right] \tag{36}$$

$$R_k(\lambda) = N(\lambda) - 2 \ln \frac{\sinh \pi\lambda/k}{\pi\lambda/k}$$

$$- 2C - 2 \sum_{n=1}^{\infty} \left[N(nk) - N\left(\sqrt{\lambda^2 + n^2 k^2}\right) \right] \tag{37}$$
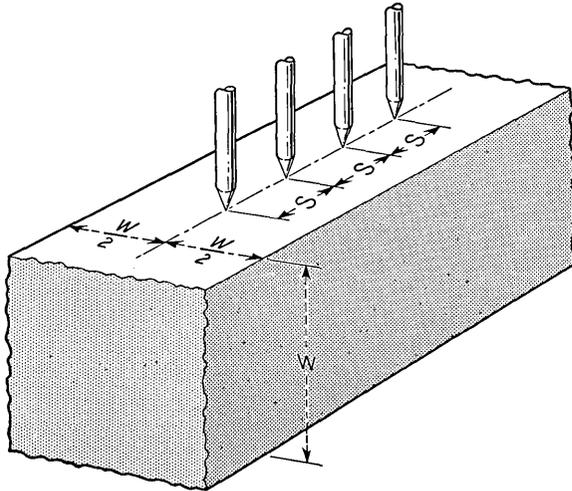
Fig. 13 — Linear probe on square filament.

Equation (36) is best for small values of λ, equation (37) for large values. In either case, one need consider only $n = 1$ to get five-place accuracy.

The potential of a plane grid of sources may be used to calculate the correction divisor for four-point resistivity measurements on filaments. The case of square filaments is illustrated in Fig. 13; the correction divisor for this case is

$$C.D. = \frac{2s}{w} \left[R_2 \left(\frac{s}{w} \right) - R_2 \left(\frac{2s}{w} \right) \right] \tag{38}$$

For example, when $s/w = \frac{2}{3}$, C.D. = 3.10. This value agrees very well with some experiments in which the reading of a four-point probe on a long germanium filament was compared with the resistivity determined from the potential gradient in the filament when current flowed through its entire length.

The usefulness of the tabulations for three-dimensional arrays of point sources was tested by calculating the Madelung constant α for the Coulomb energy of an ionic crystal of the sodium chloride type. The calculation involves summing the potentials of a number of lines of sources of alternating sign; these potentials are obtained in the same way as for the calculation of the base resistance of a point-contact transistor. Nine terms must be considered to make full use of the accuracy of the tables. It is found that

$$\alpha = 2 \ln 2 + 4 \left[N \left(\frac{1}{2} \right) - N \left(\frac{\sqrt{2}}{2} \right) - 2N(1) + 2N \left(\frac{\sqrt{5}}{2} \right) \right.$$
$$+ N \left(\frac{3}{2} \right) - 2N \left(\frac{\sqrt{10}}{2} \right) + 2N \left(\frac{\sqrt{13}}{2} \right) \tag{39}$$
$$\left. - N \left(\frac{3}{2} \sqrt{2} \right) - 4N(\sqrt{5}) \right]$$

and a value 1.74755 is obtained. The correct value is 1.747558.[3]

Plane boundaries between media of different but finite resistivities (or different dielectric constants or different thermal conductivities) can be treated by the method of images.[4] When infinite image systems arise in this sort of problem, the tabulations given in this article can be applied.

CONCLUSION

The potential of a line of point sources can be obtained from either of the functions M and N, which have been defined and tabulated. In addition to the tabulated functions, only elementary functions are required for this purpose.

The functions M and N represent the potentials of definite source systems. This fact enables one to visualize the construction of solutions to potential problems involving lines of point sources.

APPENDIX

From Fig. 1, the superposition principle, and the usual q/r potential of a point source, it may be seen that

$$M(\lambda) = 2 \sum_{n=1}^{\infty} \left(\frac{1}{n} - \frac{1}{\sqrt{n^2 + \lambda^2}} \right) \tag{40}$$

According to the binomial theorem,

$$\frac{1}{\sqrt{n^2 + \lambda^2}} = n^{-1} + \left(-\frac{1}{2} \right) n^{-3}\lambda^2 + \left(-\frac{1}{2} \right)\left(-\frac{3}{2} \right) \frac{1}{2!}\, n^{-5}\lambda^4 \cdots \tag{41}$$

so that

$$M(\lambda) = -2 \sum_{1}^{\infty} \left[\left(-\frac{1}{2} \right) n^{-3}\lambda^2 + \left(-\frac{1}{2} \right)\left(-\frac{3}{2} \right) \frac{1}{2!}\, n^{-5}\lambda^4 \cdots \right] \tag{42}$$

or

$$M(\lambda) = \zeta(3)\lambda^2 - \tfrac{3}{4}\zeta(5)\lambda^4 + \cdots , \tag{43}$$

where

$$\zeta(\sigma) = \sum_{n=1}^{\infty} n^{-\sigma} \tag{44}$$

is the Riemann zeta function. Equation (43) indicates a convenient approximation to $M(\lambda)$ for very small λ:

$$M(\lambda) \approx \zeta(3)\lambda^2 = 1.202\lambda^2 \tag{45}$$

The binomial expansion, equation (41), converges more rapidly, the higher the value of n. It is, therefore, worthwhile to split up the sum in equation (40) into two parts. The first part will be a sum of the terms from $n = 1$ to $n = m$, calculated with the original expression for the summand. The second part will be a summation from $n = m + 1$ to $n = \infty$ of the binomial expansion of the summand. It is expedient to

take $m = 1$ for $\lambda < 1$. Thus,

$$M(\lambda) = 2 - 2(1 + \lambda^2)^{-1/2} + A_3\lambda^2 - \tfrac{3}{4}A_5\lambda^4 + \tfrac{5}{8}A_7\lambda^6$$
$$- \tfrac{35}{64}A_9\lambda^8 + \tfrac{63}{128}A_{11}\lambda^{10} - \tfrac{231}{512}A_{13}\lambda^{12} + \cdots \quad (46)$$

where

$$A_\sigma = \zeta(\sigma) - 1 = \sum_{n=2}^{\infty} n^{-\sigma} \quad (47)$$

Table II gives numerical values of the coefficients of equation (46).[5] Table I is based mainly on calculations with $m = 3$. These were done at the Laboratories' IBM installation.

TABLE II — NUMERICAL VALUES OF COEFFICIENTS IN EQUATION (46)

$$M(\lambda) = 2 - 2(1 + \lambda^2)^{-1/2} + .20205\ 69032\lambda^2 - .02769\ 58163\lambda^4$$
$$+ .00521\ 82984\lambda^6 - .00109\ 83398\lambda^8 + .00024\ 32335\lambda^{10}$$
$$- .00005\ 53648\lambda^{12} + \cdots$$

TABLE III — SOME VALUES OF THE DERIVATIVE $M'(\lambda)$

λ	$M'(\lambda)$	λ	$M'(\lambda)$
0.05	0.11982	1.30	0.95001
0.10	0.23734	1.35	0.93512
0.15	0.35040	1.40	0.92004
0.20	0.45709	1.45	0.90489
0.25	0.55586	1.50	0.88975
0.30	0.64556	1.55	0.87471
0.35	0.72545	1.60	0.85982
0.40	0.79519	1.65	0.84513
0.45	0.86714	1.70	0.83068
0.50	0.90467	1.75	0.81649
0.55	0.94528	1.80	0.80259
0.60	0.97736	1.85	0.78900
0.65	1.00174	1.90	0.77568
0.70	1.01926	1.95	0.76270
0.75	1.03077	2.00	0.75003
0.80	1.03709	2.05	0.73768
0.85	1.03900	2.10	0.72564
0.90	1.03719	2.15	0.71391
0.95	1.03229	2.20	0.70249
1.00	1.02487	2.25	0.69136
1.05	1.01541	2.30	0.68053
1.10	1.00431	2.35	0.66999
1.15	0.99195	2.40	0.65972
1.20	0.97862	2.45	0.64973
1.25	0.96457	2.50	0.64000

TABLE IV — SOME VALUES OF THE DERIVATIVE $N'(\lambda)$

λ	$-N'(\lambda)$	λ	$-N'(\lambda)$
1.00	0.02487	1.55	0.00063
1.05	0.01766	1.60	0.00046
1.10	0.01257	1.65	0.00033
1.15	0.00896	1.70	0.00024
1.20	0.00640	1.75	0.00019
1.25	0.00457	1.80	0.00012
1.30	0.00326	1.85	0.00009
1.35	0.00233	1.90	0.00006
1.40	0.00167	1.95	0.00004
1.45	0.00120	2.00	0.00003
1.50	0.00086		

Although equation (46) certainly converges for all values of λ less than 2, it is not efficient for $\lambda > 1$. For $\lambda > 1$, it is better to work with the function N. According to Madelung,[6]

$$N(\lambda) = 2\pi \sum_{n=1}^{\infty} iH_0^{(1)}(i2\pi n\lambda) \qquad (48)$$

where $iH_0^{(1)}(ix)$ is the Hankel function tabulated in Jahnke & Emde.[7]

The functions $M(\lambda)$ and $N(\lambda)$ are tabulated in Table 1. $M(\lambda)$ and $N(\lambda)$ are connected by equation (4). Because of the overlapping regions of λ-values for which the expressions for M and N converge with reasonable rapidity, it is possible to tabulate both M and N for all values of λ. However, for satisfactory graphical interpolation, it is necessary to use the smaller of the two functions, so that M is given in Fig. 3 for small values of λ and N is given in Fig. 4 for large values of λ. Table III gives some values of the derivative, $M'(\lambda)$. The field E of a line of point charges can be calculated from M' in accordance with the equation obtained by differentiating equation (3) with respect to d. Thus

$$E = \frac{q}{a^2}\left[\frac{1}{\lambda^2} + M'(\lambda)\right] \qquad (49)$$

The derivative $N'(\lambda)$ is tabulated in Table IV.

1. L. B. Valdes, Resistivity measurements on germanium for transistors, Proc. I.R.E., **42**, pp. 420–427 (1954).
2. L. B. Valdes, Effect of electrode spacing on the equivalent base resistance of point-contact transistors, Proc. I.R.E., **40**, pp. 1429–1434 (1952).
3. J. Sherman, Chem. Revs., **11**, p. 93, (1932).
4. W. R. Smythe, Static and Dynamic Electricity, p. 113, McGraw-Hill, 1939.
5. Values of A_σ were obtained from the British Association for the Advancement of Science Mathematical Tables, **1**, p. xxv.
6. E. Madelung, Physikalische Zeitschrift, **19**, p. 524 (1918).
7. Jahnke & Emde, Functionentafeln, Fourth Edition, pp. 236–242.

Experiments on the Interface between Germanium and an Electrolyte

By W. H. BRATTAIN and C. G. B. GARRETT

Measurements have been made of the electrode potential of p- and n-type germanium in contact with aqueous solutions of KOH, KCl and HCl as a function of anodic and cathodic current and of incident light intensity. For anodic currents, the measured electrode potential can be separated into three parts: the reversible electrode potential corresponding to the anodic reaction, depending only on the solution; an overvoltage of the usual form; and a term $(kT/e) \ln (p_1/p)$, where p is the equilibrium hole concentration and p_1 is the concentration just inside the space-charge region of the germanium. The anodic current is determined by flow of holes to the surface, so that the current saturates for n-type germanium but not for p-type. The saturation current is determined by body and surface generation of holes and by creation of excess holes by light. There is a current gain of 1.4 to 1.8. In addition, there is a small "leakage" current not dependent on hole supply. Similar statements may be made for cathodic current, except that the electrode potential and current are determined respectively by the concentration and supply of electrons instead of holes, the current gain is of the order of unity, and the leakage current is larger. Complicating time changes were observed for cathodic but not for anodic currents. The measurements may be understood in terms of simple thermodynamic considerations, based on the idea that the anodic reaction is with holes, the cathodic reaction with electrons, in the semiconductor; the behavior for very small currents depends on a competition between the anodic and cathodic reactions, which may be treated by simple rate process considerations. A comparison is made with experiments on the germanium-gas interface by Brattain and Bardeen, to which similar considerations may apply.

A. SURVEY OF THE WORK

1. INTRODUCTION

This paper reports studies of germanium electrodes in contact with aqueous electrolytes. Measurements have been made of electrode po·

tentials for germanium electrodes in contact with aqueous solutions of potassium hydroxide, potassium chloride and hydrochloric acid. Measurements were made as a function of anode or cathode current and also of light intensity. In one series of experiments with n-type germanium simultaneous measurements were made of changes in the density of the minority carriers, i.e., holes.

The measured half cell potentials may be understood in terms of the following conclusions. At a germanium anode, the primary reaction is with holes in the semiconductor and not with electrons.[1] Because of the long lifetime of minority carriers in high purity germanium, one can establish a steady (quasi-equilibrium) state in which the density of minority carriers may differ by many orders of magnitude in either direction from the equilibrium value. One way in which this may be done is to illuminate the surface. Illumination of the surface creates equal numbers of holes and electrons. The effect upon the density of majority carriers is not very large but the relative change in the density of minority carriers may be considerable. It is found that the effect of light on the electrode potential for a p-type germanium anode is rather small, but that for n-type the effect is large. The electrode potential is found to depend upon the concentration of holes in the way which one would expect from thermodynamics if the holes are taking part as one of the components in the anodic process. From this we conclude that the primary reaction at a germanium anode is with holes. Physically the primary process in oxidation must be the transfer of the electron from the electrolyte into the valence band of the semiconductor. It has been found however that the current flowing is larger than the flow of holes by a factor of the order of 1.6. This seems to indicate that once one electron has been transferred from the electrolyte a further transfer of charge, this time into the conduction band, can take place. This gives rise to a current multiplication which one may compare with the current multiplication occurring at the collector of a point contact transistor.

At a germanium cathode the situation is complicated by time effects. In some solutions these are relatively unimportant or are confined to a particular range of electrode potential; in other solutions it is difficult to get reproducible results at all. In general it is found, however, that the primary reaction occurring on first switching on the current is with electrons rather than with holes. The evidence for this is the fact that the electrode potential for cathodic current is found initially to depend in the same way on electron concentration as does the electrode potential for anode current on hole concentration.

The experiments also give information as to the rapidity with which

minority carriers can recombine at the interface. Note that this process is quite distinct from the flow of carriers across the surface (See Appendix 1). The surface recombination velocity does not seem to vary much with the current crossing the surface, and is of the same order of magnitude as for a dry germanium surface which has been given an oxidizing etch and left in a neutral gas environment.

The change-over from anodic to cathodic process takes place near zero current and may be studied by observing the change from positive to negative surface photo-effect. From these measurements one may make tentative conclusions as to the form of the over-voltage and the equilibrium potential for the anode and cathode reactions. Unfortunately, these experiments do not give any insight as to the nature of the anodic and cathodic processes other than the part played in them by the carriers in the two energy bands in the semiconductor. Likewise no information is obtained as to the location of charge in the surface of the semiconductor. It is not possible to say how much of this charge resides in the comparatively wide space-charge region at the semiconductor surface and how much is located in surface states. The reason for this is that the quantities measured in the experiment may be related one to another by purely thermodynamical considerations.

2. HISTORY

It is an almost invariable rule that at any interface between two phases there is a separation of charge forming a space-charge double layer. This is of particular interest at a semiconductor surface because the space charge region extends relatively far (of the order of 10^{-4} cm.) into the semiconductor. The space charge region at the surface of a semiconductor is of great importance in semiconductor technology. A simple theory for the space charge region at the surface of a semiconductor was proposed some years ago by Mott[2] and by Schottky.[3] There are some features of the space-charge region which are still somewhat of a mystery, however. From the simple theory one would predict that the application of a strong electric field normal to the surface should change the conductivity of the semiconductor. This change in surface conductivity has been found, but is smaller by about an order of magnitude than one would predict.[4] This observation can be understood if one supposes that the interior is shielded by "surface states" of the type proposed by Bardeen.[5] One might suppose that the existence of these surface states could be confirmed by the type of experiment in which one measured

some sort of "surface potential", viz. a measured potential difference between two phases, one of which is the semiconductor.

One such study has been made by Brattain and Bardeen.[6] In these experiments the authors measured the contact potential and surface photo-effect for a germanium surface exposed to various gases. Now any such germanium surface must be covered by a layer of charged particles viz., ions, and such a layer of charged particles is accompanied by a field normal to the surface. One would therefore expect that by changing the gas environment, and therefore the magnitude and possibly the sign of the adsorbed ionic charge, one could alter the contact potential and the magnitude and sign of the surface photo-effect. These measurements show such changes and were interpreted in terms of a particular model of surface states. Other interpretations are however possible.[7] In view of the rather considerable complexity of the Brattain-Bardeen surface states model, one would like to have independent evidence as to its validity. Specifically, it would be advantageous to proceed further into the extremes of surface potential in order to determine the exact location in energy of the postulated surface states. The experiments to be reported in this paper were originally undertaken with this aim in view.

There is another way in which one may change the sign and magnitude of an ionic surface charge on the surface of a semiconductor. This is to make the semiconductor one electrode of an electrolytic cell. By changing the magnitude and sign of the cell current, one should be able to change the nature of the ion layer near the germanium surface. Some qualitative experiments of this type were reported by one of us (WHB) some years ago.[8] The experiments were carried out with silicon. The surface photo effect, first found with a silver chloride electrode by Becquerel,[9] was measured, and was found to depend very sharply on the current flowing. The reason for the existence of a surface photo-effect may be thought of as follows. Suppose that no net current is crossing the surface. In the dark the rates of flow of electrons and of holes to and from the surface will all be the same. When the surface is illuminated this balance is upset; the surface potential must then re-adjust itself in such a way that the net flow of electrons to the surface is equal to the net flow of holes to the surface. For a given intensity of illumination, these two rates will be the higher, the higher the surface recombination velocity. It therefore follows that a high surface recombination velocity would give rise to a low surface photo-effect and vice versa. At the time these experiments were done, a qualitative interpretation was given in terms of an analogy with a rectifier. The object of the present work was more

completely to understand the semiconductor-electrolyte system, and, if possible, to extend the experiments done on the germanium-gas system to the condition in which the change in surface potential is so large that the exact location of the surface states may be determined. As has been stated above, the latter hope has not been fulfilled.

3. PLAN OF THE PAPER

In the next paragraph an outline of the experiments is given. Most of the experiments have been carried out with blocks of germanium so large that every dimension is great in comparison with the minority carrier diffusion length. These are reported in Section B. Experiments designed to study the changes in minority carrier density are described in Section C. The conclusions to the experiments described in this paper are given in Sections D and E.

4. OUTLINE OF THE EXPERIMENT

Standard electro-chemical methods[10] have been used to measure electrode potential with light as an additional variable. The electrode potential for p-type and n-type germanium in contact with various solutions has been measured in light of a series of known intensities.

The surface photo effect is defined by the quantity $(dV^E/dL)_{IE}$, where L is defined as the number of hole-electron pairs created per second per unit area, multiplied by e, the electronic charge. This is the maximum photo-current which would pass if every minority carrier were collected (See Appendix 1). This quantity may be derived from the dc measurements; it is however convenient to make an independent measure of it on an ac basis. This measurement will be reliable provided that there is no dispersion between zero frequency and the frequency used for the ac experiment.

It soon became apparent that p-type and n-type germanium behave very differently; n-type germanium shows a large positive surface photovoltage for anodic current and a small negative photo-voltage when the current is cathodic; for p-type germanium the sign of the photo-voltage is correlated with the sense of the current in the same way, but the relative magnitudes are reversed. The measurements on n-type germanium showed a sharp saturation of anodic current, the value at which saturation occurred depending on the intensity of illumination of the surface. A similar saturation was found in cathodic current for p-type germanium, although in this case the saturation was less sharp. It was suspected that these effects might be due to the exhaustion of minority

carriers. It therefore became important to have some independent means for following changes in the minority carrier density just inside the space charge region.

Fortunately, the minority carrier diffusion length in germanium is quite long. It is therefore possible to make electrochemical measurements on one side of a thin slice of germanium and measurements of minority carrier density on the other. Experiments of this type have been carried out with n-type slices of the order of 10^{-2} cm. in thickness — a distance large in comparison with the thickness of the space charge region but still small in comparison with a minority carrier diffusion length. The back surface is covered with a large-area rectifying junction. From measurements of the floating potential across this junction one calculates the hole density near the back surface. The value of the hole density near the electrolyte surface may then be deduced. Since the slice is thin in comparison with a diffusion length, these two densities will not usually be very different.

The rectifying junction at the back of the thin slice has another use. As an alternative to creating surplus holes by shining light on the electrolyte surface, it is also possible to inject them by passing a positive current from the alloyed region into the semiconductor. This gives us a way of calibrating the intensity of the light. Thus one can measure the floating potential at the rectifying contact, the electrode potential and the ac surface photo effect as functions of electrolyte current, current across the rectifying barrier and light intensity. These experiments have shown that, at anode current I^E, the electrode potential V^E is given by the following equation:*

---

* The sign convention used in this paper for electrode current and electrode potential is explained in the table given below:

Voltage		Current	
Plus	Minus	Plus	Minus
Less noble	More noble	Cathodic	Anodic

The convention for currents agrees with that common in semiconductor physics (conventional current *into* the semiconductor positive); the convention for voltages agrees with that of the American Chemical Society, but disagrees with that of the American Electrochemical Society and of European electrochemists. Electrode voltages are given with respect to the particular standard electrode used by us (silver-silver oxide in N/10 KOH); to convert to the normal hydrogen electrode, subtract 0.41 volts.

The words "cathodic" and "anodic" also require explanation. That sense of current is called anodic which tends to promote an oxidative reaction at the electrode surface; the other sense is cathodic. At and near zero current, however, both oxidative and reductive processes will be going on at the time; thus, if we use the terms anodic and cathodic to distinguish oxidative and reductive *reactions* as well as to describe the two senses of external current, we have the paradox that both anodic and cathodic processes are occurring at all times, whether the external *current* is anodic or cathodic — though one process will greatly predominate whenever the net current is at all large. (See Section D.)

$$V^E = V_{0A}{}^E - \frac{K_A}{\beta} \, \ell n \left(1 + \frac{|I^E|}{I_{0A}{}^E} \right) + \frac{1}{\beta} \, \ell n \, \frac{p_1}{p} + I^E \, R_{\text{spr}}. \qquad (1)$$

where $\beta = kT/e = 39$ volts^{-1}.

In this equation the first term is an equilibrium potential $V_{0A}{}^E$ for the anode, the second term is the ordinary Tafel[11] over-voltage; the third term shows how the electrode potential depends upon the concentration of the minority carriers and the fourth term represents the IR drop through any series resistance, such as the ohmic resistance of the germanium slice. The term which is of interest to us is the third. This may be regarded as part of the equilibrium* electrode potential, and shows that holes are one of the components in the primary anodic reaction.

If this is so, one would expect the anode current to depend upon the flow of holes up to the semiconductor surface. Let I_s be the saturation current for minority carriers, that is, the maximum rate at which they may come up to the surface from generation within the body or at the surface itself. The flow of holes up to the surface may then be related[12] to the concentration of holes just within the space charge region and the intensity of illumination by the following equation:

$$I_p = I_s \left(\frac{p_1}{p} - 1 \right) - L \qquad (2)$$

which is analogous to a similar expression for a p-n junction. The experiments show that I^E and I_p can be related by writing:

$$I^E = \alpha I_p + \Delta \qquad (3)$$

where α, the current gain for the surface is a constant and Δ is a small "leakage" current,[13] a function only of V^E.

One may check equations (1), (2) and (3) from the measurements of surface photo-effect. From these equations one finds

$$\left(\frac{\partial L}{\partial V^E} \right)_{I^E} = \beta I_s \left(\frac{p_1}{p} \right) + \frac{1}{\alpha} \frac{d\Delta}{dV^E} \qquad (4)$$

so that, for constant electrolyte current one should have, neglecting the second term of equation (4),

$$V^E = \frac{1}{\beta} \, \ell n \left(\frac{\partial L}{\partial V^E} \right)_{I^E} + \text{const.} \qquad (5)$$

---

* By imagining the rates of the various recombination processes to be vanishingly small, one may consider the disturbed (e.g., illuminated) condition of the germanium as a quasi-equilibrium state.

These measurements also enable one to find the quantity I_s, by determining the value of $(\partial L/\partial V^E)_{I^E}$ for which $p_1 = p$. There are various ways in which this may be done.

For p-type material, measurements have not yet been made with the thin slice technique. The information obtained from the experiments on the n-type slices, however, enables one to understand the experiments made on large blocks of either type. In this way, one may check equation (5) on p-type material in the cathode direction and so show that, by analogy with equation (1), the initial electrode potential for p-type material in the cathode direction is given by the equation:

$$V^E = V_{0c}{}^E + \frac{K_c}{\beta} \, \ell n \left(1 + \frac{I^E}{I_{0c}{}^E} \right) - \frac{1}{\beta} \, \ell n \frac{n_1}{n} + I^E R_{\mathrm{spr}}. \qquad (6)$$

Further, the current flowing is related to the light current L and the *electron* density n_1 just inside the space-charge region by the equations

$$\left. \begin{aligned} -I_n &= I_s \left(\frac{n_1}{n} - 1 \right) - L \\ I^E &= \alpha I_n + \Delta \end{aligned} \right\} \qquad (7)$$

which are analogous to equations (2) and (3). The current gain α is different in the anode and cathode directions; the same is true for the dependence of the leakage current Δ on V^E. Otherwise the anodic properties of n-type germanium and the initial cathodic properties of p-type germanium are identical.

These considerations account for the behavior when the current is of that sense which gives a large photo-voltage. We shall show further that equations (1) and (6) describe also the small positive photo-voltage found for p-type germanium in the anode direction and for the small negative photo-voltage in the cathode direction on n-type. Holes are required for the anode process; it is more difficult to change the concentration of holes in p-type material than in n-type material, and therefore the measured electrode potential is less sensitive to light. A similar argument applies for electrons in n-type material. The change in phase of surface photo-voltage thus corresponds, qualitatively, to the changeover from anodic to cathodic processes. We shall return to this question in Section D. For the moment we are chiefly interested in the surface photo-effect as a means for determining minority carrier density for n-type material in the anode direction and for p-type material in the cathode direction.

B. Experiments on Large Blocks

1. PREPARATION OF SAMPLES

The samples were cut from single crystals of germanium grown in such a manner as to have reasonably uniform resistivity and long lifetime. The resistivity in each case was determined by making measurements on a bar cut from an adjacent part of the crystal. The body lifetime is estimated as follows. All of the surface of the sample is sand-blasted except for one small area. On this area a phosphor bronze point is set and formed. The sample surface is illuminated with a narrow slit of light a known distance away from the point; the rate of collection of minority carriers by the point is then determined as a function of the distance between the point and the line of the light. The body lifetime may then be estimated by standard theory[14] assuming that the surface recombination velocity is large. The samples were about $1''$ square and $\frac{1}{2}''$ thick.

In order to make good ohmic contact to the sample, one face was plated with rhodium. The opposite face was then lapped or sand-blasted and given a two minute etch in CP4. The sides of the samples were covered with an insulating layer of Ucilon lacquer.

2. APPARATUS

The apparatus used for the electrochemical measurements is shown in Fig. 1. The left-hand glass vessel contains the solution to be used. Into this is set the germanium sample, with the etched surface exposed to the electrolyte. The glass vessel also contains a loop of platinum wire for passing current. The germanium sample is held in position by setting it against a copper tube connected to a vacuum line: the copper tube also serves as a good electrical contact to the rhodium plating. The liquid in the germanium half-cell was stirred by means of a paddle. Electrical connection to the reference half-cell in the right-hand glass vessel is made through a salt-bridge. The reference half-cell used throughout the measurements consisted of silver-(silver oxide) in deci-normal potassium hydroxide solution. This half-cell has a potential of -0.41 volts against the standard hydrogen electrode where the symbol "—" means "more noble".

Arrangements were made, as shown in Fig. 1, to illuminate the germanium surface. Two sources of light were used. One, for steady illumination, consisted of an American Optical Company microscope illuminator. The intensity of the light from this source could be altered by

setting the iris diaphragm. For determining the surface photo effect on an ac basis, a second source, the light from which could be chopped at a convenient low frequency (about 40 cyc/sec), was used. The intensity of the chopped light could be controlled by choosing one of a series of fixed diaphragms.

The electrical circuit for the measurements is shown in Fig. 2. Dc measurements of electrode potential were made on a vacuum tube voltmeter. Measurements of the ac surface photo-voltage were made on a wave analyzer tuned to the fundamental of the chopper frequency. Each of these instruments had an input impedance of about 2 megohms, so that measurements were effectively made on open circuit. Anodic or cathodic current could be passed across the germanium half-cell by adjusting the setting of the current source, the internal impedance of which was also of the order of megohms. Arrangements were made so that the current could be switched in suddenly and switched out again after a short time. In order to avoid the accumulation of the product of electrochemical reaction at the germanium surface, measurements were taken by switching the current alternately in the anode and cathode direction. Such measurements were found to be sufficiently reproducible. Readings could be taken in a time of the order of half a second after switching on the bias current. In cases where there were changes with time, the values obtained immediately after switching on the current were adopted.
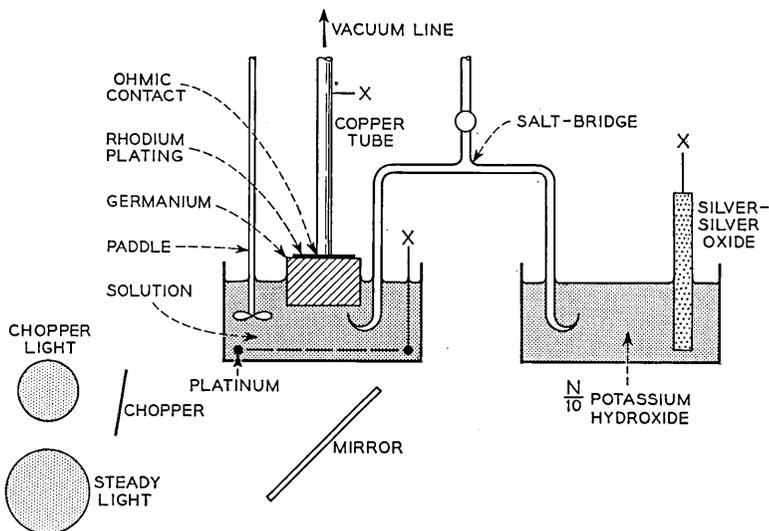

Fig. 1 — Experimental arrangement of apparatus.
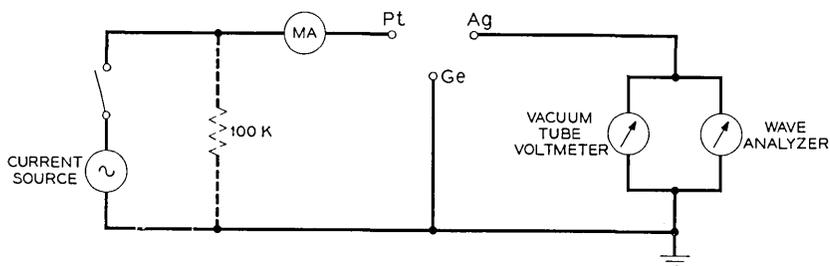
Fig. 2 — Electrical circuit for measurements of $(\partial V/\partial \alpha L)_I$ and V^E as functions of I^E and L.

3. PROCEDURE

3.1. *Light Calibration*

The relative light intensities corresponding to the different diaphragms in the light chopper were determined by measurements made with the thin slice technique (Section C). The absolute values of the light intensities may be estimated by taking these measurements and multiplying by the ratio of areas. There is some uncertainty involved in the latter process because, although attempts were made to insure that the intensity or illumination on the germanium surface was as uniform as possible, some non-uniformity remains. It may be assumed that the relative light intensities were known to about 5 per cent and the absolute magnitude to perhaps 25 per cent.

3.2. *Measurements of Electrode Potential and Surface Photo Voltage*

These could be started as soon as the freshly etched germanium sample was set into the electrolyte. It was however found that, during the first few minutes, there was a small but significant drift in these measurements. Sufficient time was therefore allowed to bring the surface into a steady-state condition. This process could be accelerated by switching a current of about 1 mA alternately in the anode and cathode sense several times.

Measurements were made as follows. The current desired could be set on the current source and switched into the germanium half-cell for a time sufficient to make the measurement of electrode potential and surface photo voltage. At each current, the chopper diaphragm was chosen in such a way that the surface photo voltage was about 5 millivolts, that is, small in comparison with kT/e, but large enough to be

measured accurately. Measurements were made in this way at a series of values of anode and cathode current.

The steady light was now switched on. The intensity of the steady light could be set equal to that produced by a standard chopper diaphragm in the following manner. The chopper was stopped and a standard diaphragm inserted. A suitable current was now passed through the electrolyte to bring the surface into a condition in which the measured electrode potential was extremely sensitive to light intensity. The iris diaphragm of the steady light was then adjusted until the steady light gave the same electrode potential as did the light from the standard chopper diaphragm. With the surface now maintained in this steady illumination, the set of measurements of electrode potential and ac surface photo voltage was repeated exactly as before. After completion of this set of measurements the iris diaphragm on the steady light was readjusted to such a setting that the new intensity of the steady light was equal to the sum of the previous steady light intensity and that given by the standard chopper diaphragm, i.e., twice that given by the standard chopper diaphragm. The measurements of electrode potential and surface photo-voltage as a function of anode and cathode current were then repeated. In this way a series of such measurements were made at light intensities equal to integral multiples of that given by the standard chopper diaphragm.

3.3. *To Verify Equation (5) Directly*

At each value of the electrolyte current, a series of readings of electrode potential and surface photo voltage was made, varying the setting of the steady light iris diaphragm. No attempt was made to estimate the intensity of the steady light at each point. The experiments were repeated at a series of anode currents for n-type material and at a series of cathode currents for p-type material.

3.4. *Direct Determination of the Over-Voltage Curve at Constant Minority Carrier Density*

As each value of electrolyte current, the steady light was adjusted to give some standard photo voltage. The electrode potential was then read and the process repeated at some other electrolyte current. Measurements made on n-type material in the anode direction in this way may then be compared with the p-type anode over voltage, and measurements made on p-type material in the cathode direction in this way with the n-type cathode over voltage.
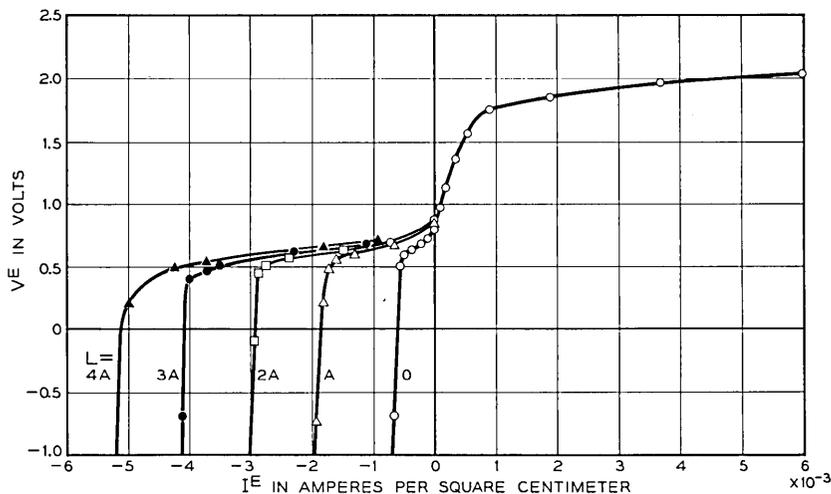
Fig. 3 — Plot of V^E against I^E for n-type in $N/10$ KOH, at different light intensities.

4. RESULTS

4.1. *n-Type in Decinormal Potassium Hydroxide*

$\rho = 15$ ohm cm. $\tau > 4000$ μsec. $T = 27°C$ area $= 5.4$ cm^2

Fig. 3 shows the observed relation between electrode potential and current in various steady lights. The lights used were multiples of the intensity A obtained with the $1\frac{1}{2}''$ diaphragm in the light chopper. It will be seen that in each case the current saturates very abruptly at some particular value in the anode direction. The saturation is not perfect, however. This is shown in Fig. 4, which shows measurements made in the dark extended to greater anode potentials. This graph shows that at sufficiently high anode potential extra current will pass even though the supply of minority carriers has been used up.

From Fig. 3 we may plot the anodic current flowing at some fixed potential, say $V^E = 0$, as a function of light intensity (Fig. 5). From equations (2) and (3) it follows that the slope of the curve should be αL. The value of αL for the standard light intensity is therefore found to be

$$\alpha A = 1.13 \times 10^{-3} \text{ amp/cm}^2$$

Postponing for the moment the question of the value of α, we now
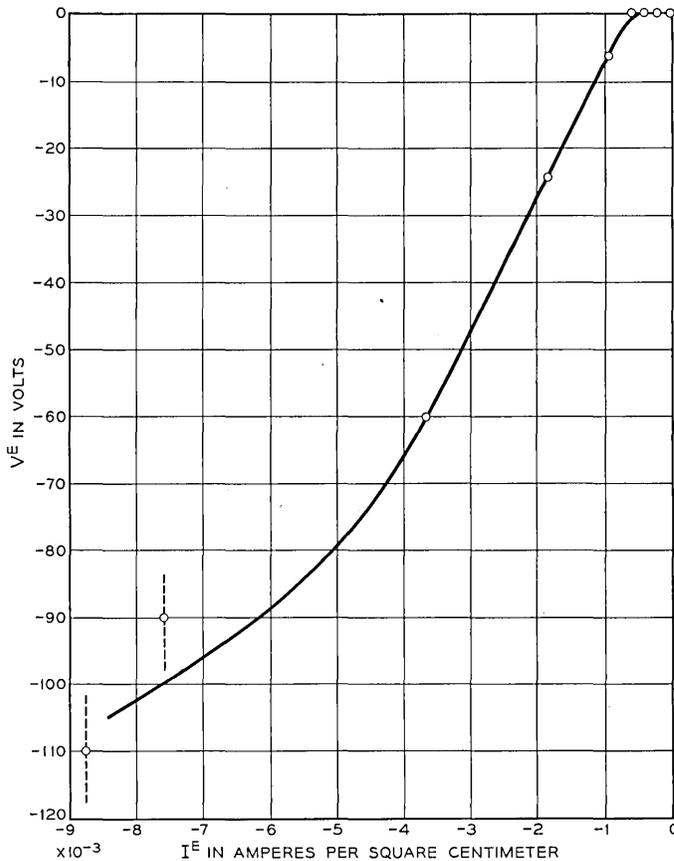
Fig. 4 — Plot of V^E against I^E for n-type, with no steady light, continued to large negative values of V^E.

know the value of αL for each diaphragm of the light chopper. From each reading of photo voltage therefore, it is possible to evaluate the quantity $(\partial V^E/\partial \alpha L)_{I^E}$. Fig. 6 shows this quantity plotted semi-logarithmically against electrode potential. The left-hand branch of the curve corresponds to a positive photo voltage; that is to say, the electrode potential increases on shining light on the surface. At the zero in photo voltage the sign of the photo voltage changes, the right-hand branch corresponding to a negative photo voltage. In zero light, the shoulder near $V^E = 0.8$ volts is rather indeterminate and depends on the exact condition of the surface. The dashed curve in Fig. 6 represents a second set of measurements taken on a subsequent day. This graph should be

compared with the curves showing surface photo effect as a function of contact potential, as found by Brattain and Bardeen in the germanium-gas experiments. The qualitative similarity is quite striking. The fact that the surface photo voltage in the left-hand branch eventually becomes very much larger than was found in the germanium-gas experiments is due to the depletion of holes caused by passing anode current across the germanium surface (Section A). If there were perfect saturation in the anode direction, one would expect that eventually the surface photo effect should become infinitely large. It will be seen that this is not so. The eventual limiting value, which is due to the second term in the right-hand side of equation (4), may be checked with the slope of Fig. 4.

The remainder of this section will be confined to a discussion of the surface photo voltage in the left-hand branch of the curve, where one is reasonably certain that equation (4) is satisfied. The cathode direction and the region of change in sign in surface photo voltage will be discussed in Section D. From equation (2), (3) and (4) one would predict that on plotting $\beta^{-1} (\partial\alpha L/\partial V^E)_{I^E}$ against I^E one should obtain straight lines of slope unity. This is shown in Fig. 7, in which the experimental points are compared with straight lines drawn with this slope.

Fig. 8 shows $[(\partial V^E/\partial\alpha L)_{I^E}]$ plotted logarithmically against V^E at a series of anode currents. From equation (5) the slope of these curves
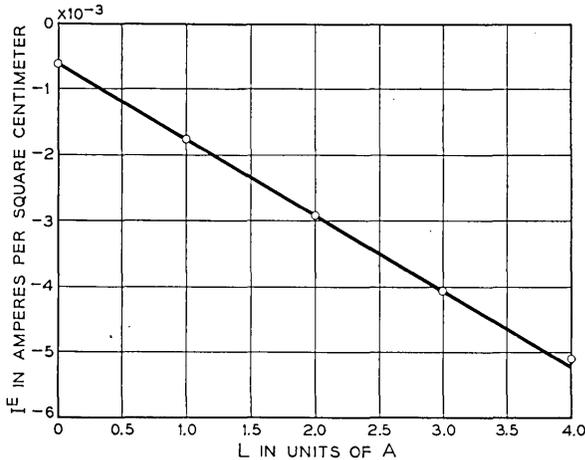
Fig. 5 — Plot of I^E at $V^E = 0$ for n-type as a function of light intensity in units of A.
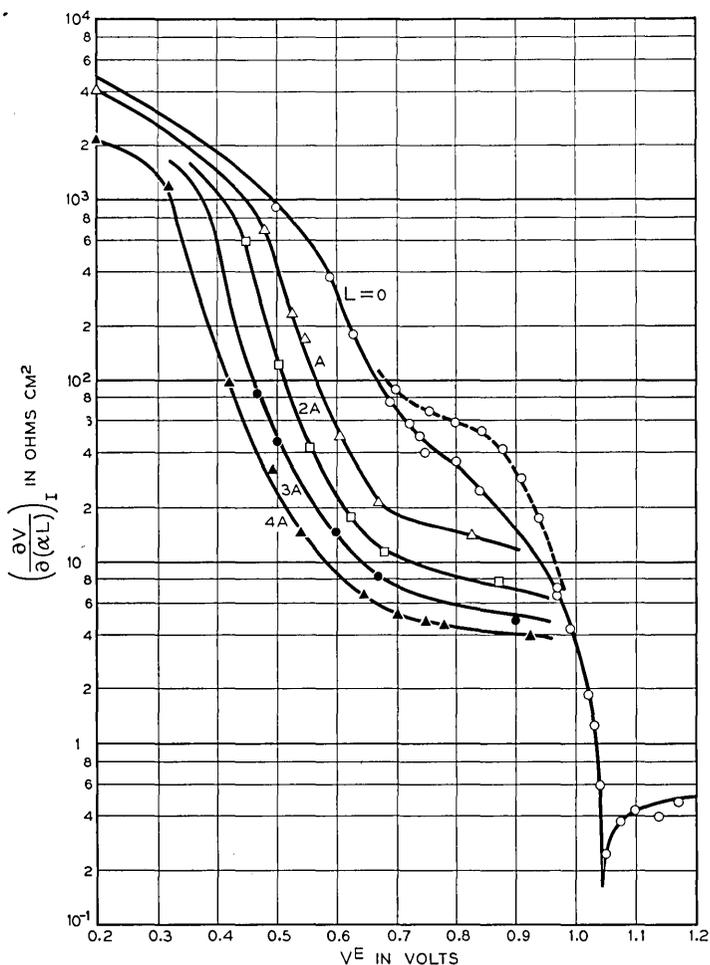
Fig. 6 — Plot of $(\partial V/\partial \alpha L)_I$ against electrode potential for n-type at various values of steady light.

should be β; the lines drawn have a slope of 37 volts^{-1}, which is slightly less than the theoretical value of 39 volts^{-1}.

The horizontal displacement of the straight lines in Fig. 8, and the relation between V^E and I^E found for p-type material for anodic currents (see below), determine the Tafel over-voltage term in equation (1). Fig. 9 shows I^E plotted semilogarithmically against (i) V^E for p-type material; and (ii) the values of V^E at the intercepts in Fig. 8 of the con-

stant-current lines with the abscissa that gives the best fit with the p-type data, after allowing for IR drop in both cases. From equation (1), this abscissa must correspond with $(p_1/p) = 1$. Using this deduction, we may translate the vertical scale of Fig. 8 into a measure of (p_1/p), see equation (4), as has been indicated on the right-hand side of that diagram. From the slope of Fig. 9, we have $K_A = 1.6$.

Now knowing the value of $(\partial \alpha L/\partial V^E)_I$ corresponding to $(p_1/p) = 1$, we may deduce from equation (4) the quantity αI_s, which is found to be 5.2×10^{-4} amps/cm^2. Another estimate is obtained from Fig. 7, from the intercept on the vertical axis corresponding to $I^E = L = 0$.
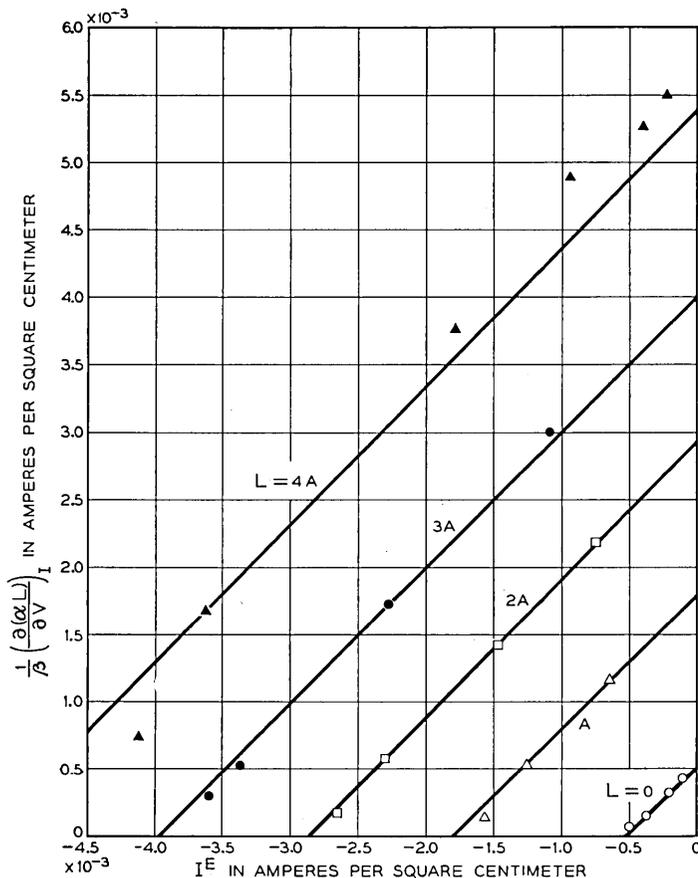
Fig. 7 — Plot of $\beta^{-1} (\partial \alpha L/\partial V)_I$ against I^E for n-type, at various values of light intensity.
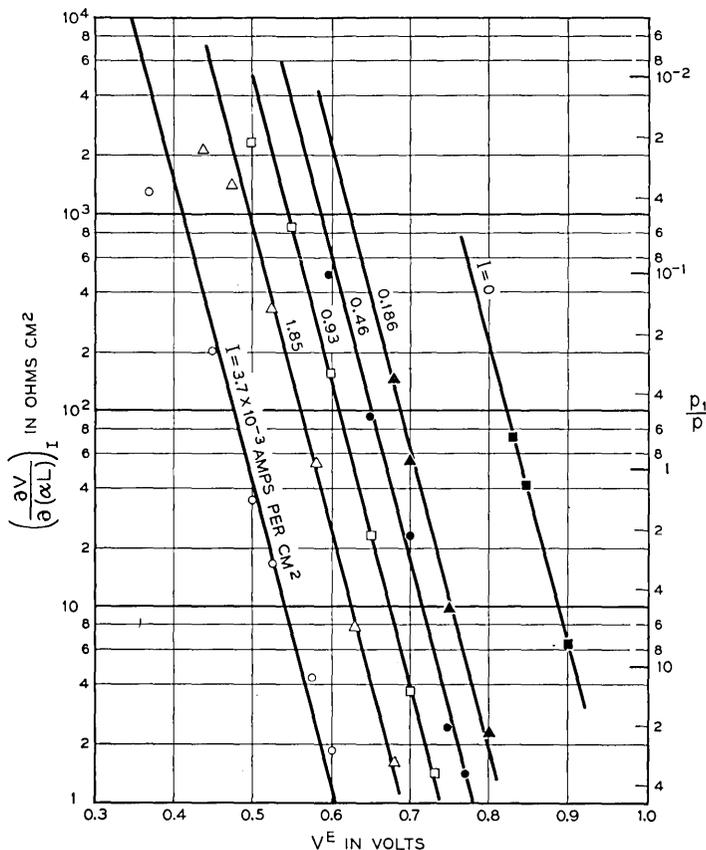
Fig. 8 — Plot of $(\partial V/\partial \alpha L)_I$ against electrode potential for n-type, at various values of anodic current. The scale on the right-hand side gives corresponding values of p_1/p.

This again gives $\alpha I_s = 5.2 \times 10^{-4}$ amps/cm². Both these estimates depend on the assumption that the second term in equation (4) is negligible: the evidence for this is that the curves in Fig. 8 are still substantially straight for some way above the abscissa corresponding to $(p_1/p) = 1$. A third estimate can be obtained by neglecting the leakage current Δ altogether, in which case αI_s is given by the intercept in Fig. 5 at $L = 0$. This gives 6.2×10^{-4} amps/cm². This value, as would be expected, is rather larger than the more accurate values obtained above.

In order to estimate α, one needs to know the magnitude of the standard light intensity. As mentioned earlier, this is difficult to do very ac-
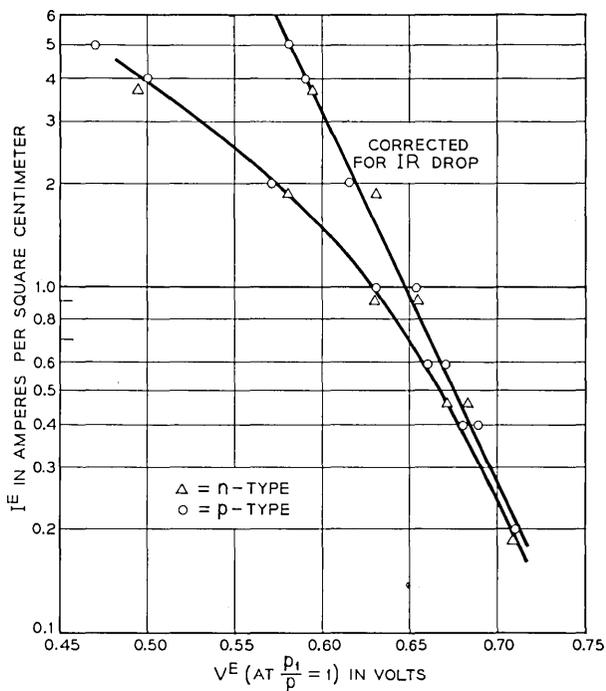
Fig. 9 — Anodic Tafel relation for both p and n type, allowing for ohmic drops in the circuit.

curately. A rough calibration, made using a photo-cell similar to the assembly used in the work of Section C, gave

$$A = 7.3 \times 10^{-4} \text{ amps/cm}^2$$

with this value of L, $\alpha = 1.6$. A better estimate of α for this material, obtained in the experiments described in the next section, is 1.80, in reasonable agreement with the above. Using $\alpha = 1.80$ and $\alpha I_s = 5.2 \times 10^{-4}$ amps/cm^2, we deduce

$$I_s = 2.9 \times 10^{-4} \text{ amps/cm}^2$$

Contributions to I_s come from thermal generation of minority carriers in the body, at the sides, and at the electrolyte surface itself. From the work in Appendix 1 it may be shown that, for a cylindrical block of radius r_0 that is thick in comparison with a diffusion length, the saturation current should be equal to $e p_0 \, V$, where

$$V = v_s + [D(1/\tau + 2v^*/r_0)]^{1/2}$$

where τ is the body lifetime, v^* is the surface recombination velocity at the sides of the cylinder, and v_s is that at the electrolyte surface itself. Using the experimental value of I_s and the value of p_0 deduced from the resistivity measurements, we find $V = 330$ cm/sec. In view of the facts that only a lower limit for τ is available, and v^* is not known at all, we can only make the crudest estimates of the various contributions to V. The experimental value would be obtained, for example, by setting $\tau = 4 \times 10^{-3}$ secs. and $v^* = v_s = 180$ cm/sec. The main conclusion to be drawn is that, even if V were entirely due to surface recombination at the electrolyte surface (which is certainly not the case), the surface recombination velocity at that surface would not be larger than that for a dry etched surface by a very considerable factor.

4.2. p-Type Germanium in Decinormal Potassium Hydroxide

$\rho = 7.5$ ohm cm $\tau > 1000$ μsecs. $T = 27°C$ area = 5.6 cm^2

The electrode potential V^E as a function of electrolyte current I^E and in various light intensities is shown in Fig. 10. (The light unit A was not quite the same as in the experiment on the n-type sample.) It will be noted that the saturation for p-type material in the cathode direction is less sharp than for n-type in the anode direction. In the cath-
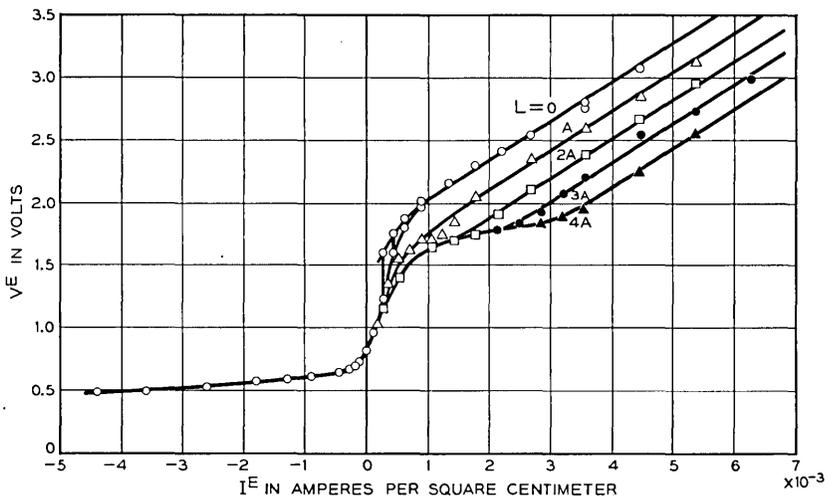
Fig. 10 — Plot of V^E against I^E for p-type in $N/10$ KOH, at different light intensities.
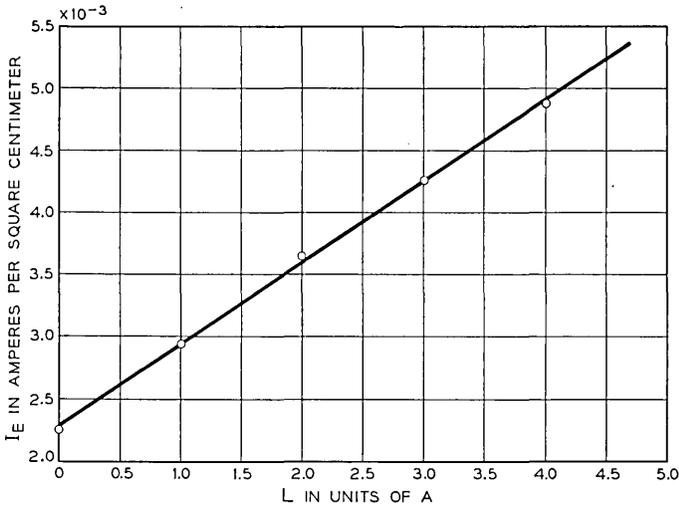
Fig. 11 — Plot of I^E at V^E = 2.40 volts for p-type as a function of light intensity in units of A.

ode direction there were two distinct regions of stable electrode potential: that for $V^E < 1.2$ volts, and that for $V^E > 1.8$ volts; for intermediate voltages, the readings were not steady, as might be expected if, in this region, there was a change-over from one electrode reaction to another. We have to consider the two regions separately. In Fig. 11, we have chosen $V^E = 2.4$ volts as a convenient value of voltage, and plotted I^E as a function of light intensity. From the slope, one deduces $\alpha L = 8.5 \times 10^{-4}$ amps/cm^2 for the unit of light intensity. A direct comparison of the α's for the n and p-type samples, using the same intensity of illumination, gave $\alpha_A/\alpha_C = 1.93$. Thus, if $\alpha_A = 1.8$, $\alpha_C = 0.93$. Within the accuracy of the experiment, there is thus no current multiplication at a p-type cathode.

Fig. 12 shows $(\partial V^E/\partial \alpha L)_I$* as a function of V^E. It will be seen that this diagram is just the converse of Fig. 6: the smaller limiting value of the surface photo voltage is now in the anode direction, and the larger, which is lower than for the n-type sample because of the less perfect saturation, is in the cathode direction. (The limiting value may again be checked with the slopes of the saturation lines in Fig. 10.) In Fig. 13, the quantity $\beta^{-1} (\partial \alpha L/\partial V^E)_I$ is plotted against I^E. The straight lines are in the predicted positions, using the theoretical value of β.

---

* For the balance of this section, the symbol α means α_C .
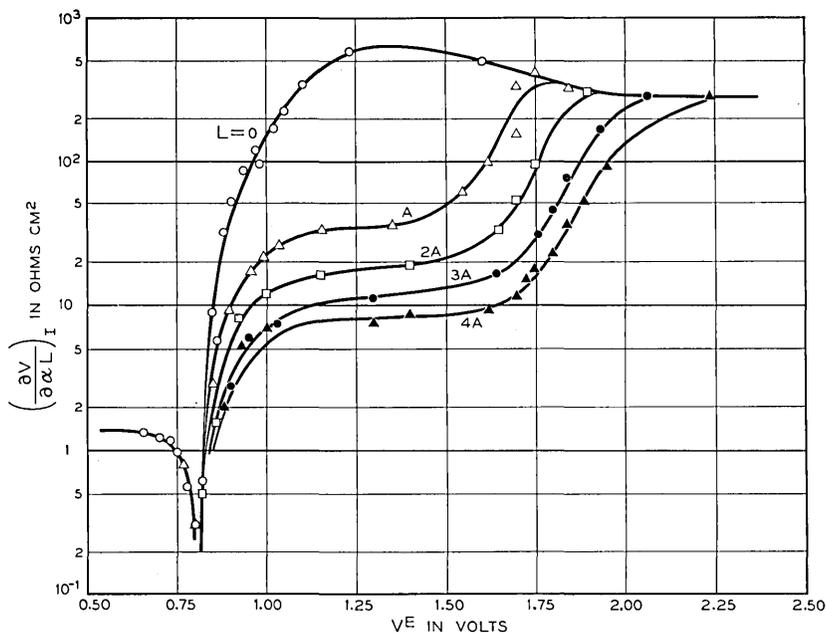
Fig. 12 — Plot of $(\partial V/\partial \alpha L)_I$ against electrode potential for p-type near the change in sign of the photo-effect, at various values of steady light.

Fig. 14 is the analogue of Fig. 8: $(\partial V^E/\partial \alpha L)_I$ is plotted logarithmically against V^E for a selection of cathode currents. The lines are drawn with their theoretical slopes. Agreement is again adequate, although not so good as in Fig. 8. The poorer saturation is again obvious in this figure, in that the experimental curves quickly depart from linearity in the section of the figure corresponding to exhaustion of electrons.

Fig. 15 has been constructed in the same way as Fig. 9, and contains points both for p-type and for n-type germanium. From the slope, the Tafel coefficient K_C *in this region* appears to be 4.3, which seems rather large. Notice that the Tafel coefficient for the lower cathode region may be quite different. (See Section D.) In constructing Fig. 15, we are able to deduce, in the same way as for n-type material, the value of $(\partial V/\partial \alpha L)_I$ corresponding to $(n_1/n) = 1$, although in this case the result is much less reliable. The resulting translation of the vertical scale of Fig. 15 into (n_1/n) is shown on the right-hand side of that figure. The corresponding value of αI_s is about 5×10^{-4} amps/cm². It is difficult to check this, as we did for the n-type measurements. If the values of $(\partial \alpha L/\partial V^E)_I$ taken in zero steady light are plotted against I^E (see Fig. 13), the resulting

curve does not even approximate to a straight line: at lower currents, the situation is complicated by the changeover to the anode reaction (see Section D), and at higher currents, by the imperfect saturation and instability of V^E. If, nevertheless, we extrapolate this curve to zero current by using the theoretical slope, we get $\alpha I_s = 2.7 \times 10^{-4}$ amps/cm^2. Under the circumstances, this does not agree too badly with the value obtained by comparing the n-type and p-type Tafel curves. We conclude that the electron saturation current for the p-type sample is of the same order of magnitude as the hole saturation current for the n-type sample. This means that the surface recombination velocities are not too different. We discuss the ratio of surface recombination velocities in the two current directions in Section D. Notice, however, that the quantity I_s is in any case significant only when (n_1/n) is neither 0 nor 1; for the case of almost complete electron exhaustion, corresponding to the higher points in Fig. 10, it is immaterial whether we ascribe the extra current flowing to the leakage current Δ or to increasing I_s. On the grounds of convenience we prefer the former description.
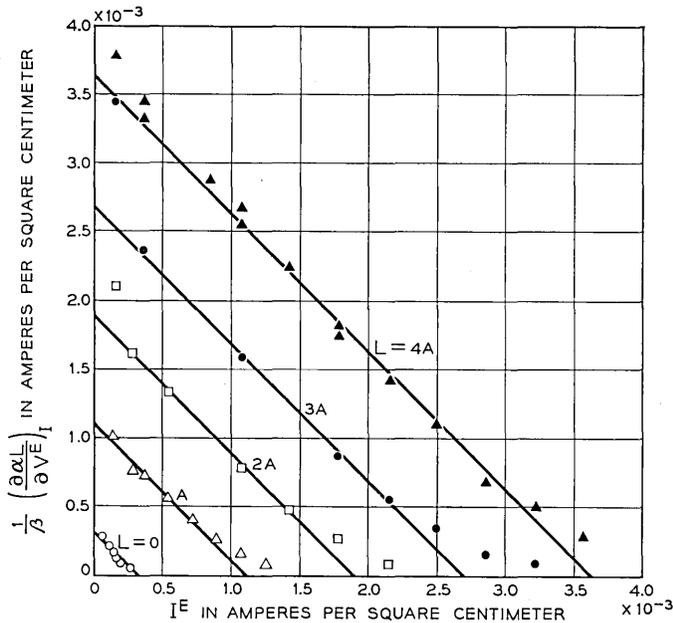
Fig. 13 — Plot of $\beta^{-1} (\partial\alpha L/\partial V)_I$ against I^E for p-type at various values of light intensity.
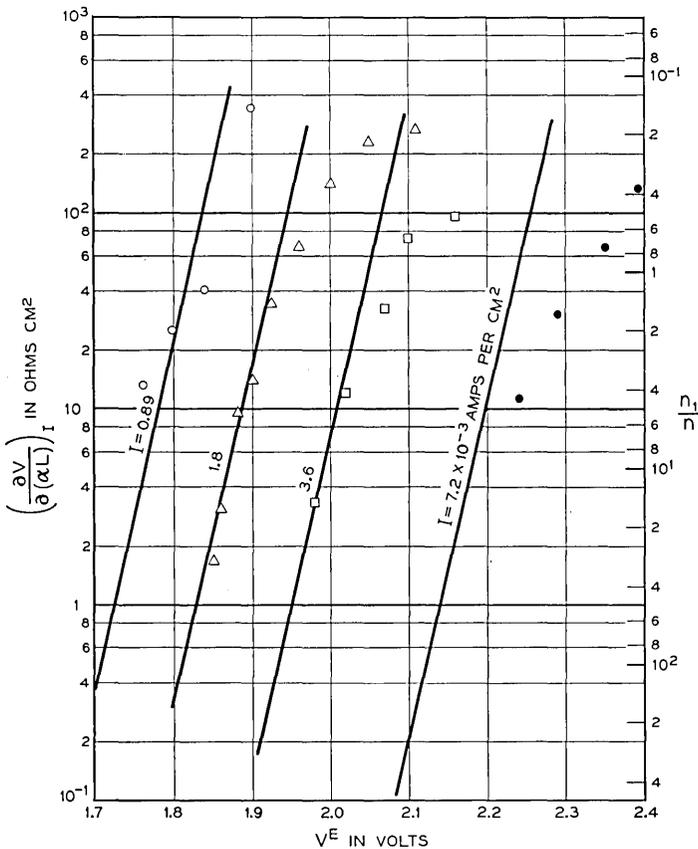
Fig. 14 — Plot of $(\partial V/\partial \alpha L)_I$ against electrode potential for p-type, at various values of cathodic current. The scale on the RHS gives corresponding values of n_1/n.

4.3. *Experiments in Other Electrolytes*

Experiments have also been carried out in solutions of potassium chloride and hydrochloric acid. Fig. 16 shows the relation between electrode potential and current in these three electrolytes, the effects of depletion of electrons or holes having been removed from the curves. Notice that the potassium chloride solution behaves like hydrochloric acid in the anode direction and like potassium hydroxide in the cathode direction. This may be because in the case of potassium chloride, the pH of the solution in the vicinity of the electrode is extremely dependent on the progress of the electrode reactions. The extreme difference be-
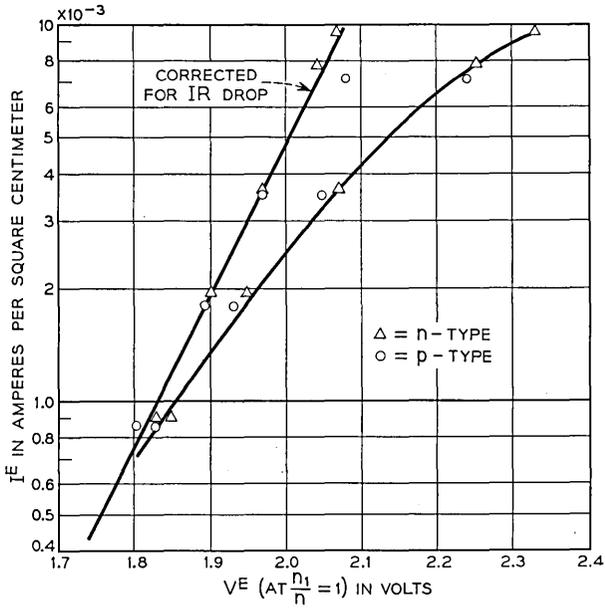
Fig. 15 — Cathodic Tafel relation for both p and n type, allowing for ohmic drops in the circuit.
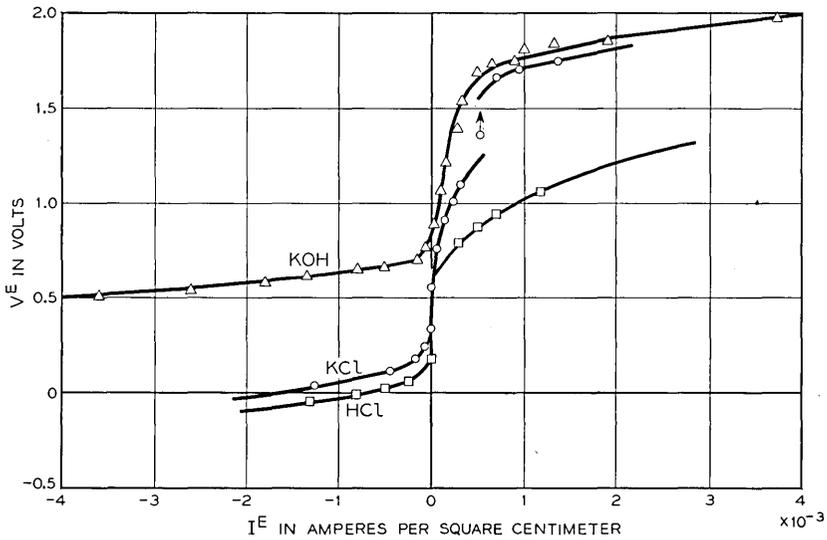
Fig. 16 — Plot of V^E, for equilibrium minority carrier density, as a function of I^E, in various electrolytes.

tween the potassium hydroxide and hydrochloric acid curves, which are roughly parallel, is about 700 millivolts, which is roughly equal to the change in the hydrogen electrode potential in going from pH1 to pH13. Differences between the three solutions were, however, observed. Specifically it was found that, whereas in potassium hydroxide cathode potentials higher than 1.80 volts were stable, no such stability was found in hydrochloric acid. Potassium chloride represented an intermediate case, behaving for the first minute or so like potassium hydroxide and then abruptly changing over to the hydrochloric acid behavior. For both hydrochloric acid and potassium chloride with p-type material it was found that, at the same time as the electrode potential was changing, the initially large photo-voltage decreased with time, went through zero, and was replaced by a small photo-voltage of the opposite sign. This one must understand as follows. The primary cathode reaction, occurring in the first few seconds, is with electrons, and, because the electrode reaction tends to deplete electrons from the p-type material, a large barrier can be built up. After this reaction has proceeded for some time, its place is taken by another reaction involving holes which are plentiful in the p-type material. This results in a change in the observed electrode potential and a change in sign of the photo voltage.

C. Thin Slices

1. EXPERIMENTS ON THIN SLICES

As mentioned earlier, it is possible, by making measurements on a thin slice of n-type germanium with a rectifying junction on the far side, to follow changes in hole density as these occur. The hole density near this rectifying junction is in fact determined by measuring the floating potential V_f. Measurements of this type enable one to demonstrate directly that equation (1) is true for anode currents irrespective of the validity of equation (2) and (3). Experiments with this geometry also enable one to determine the current multiplication factor α directly. Extra holes in the n-type region may be created in two ways: by shining light on the sample or by passing current across the rectifying junction in the forward direction. One may therefore calibrate the light intensity directly, by comparing the changes in the electrode properties produced by a given light with the changes produced by passing a given current across the rectifying junction.

The further analysis of these experiments however is complicated. The geometry is much less nearly one-dimensional, because of recombi-

nation at the sides, in the body and at one or both surfaces. Not all of the minority carriers injected from one side will reach the other. Also, the series resistance in this geometry is necessarily much larger than with a large block. In any case, the analysis duplicates to a large extent the experiments described in Section B. We therefore confine ourselves to describing such experiments as cannot be done with the large block technique.

2. MATERIAL AND GEOMETRY

The samples used consisted of M-1777 p-n-p power transistors, without collectors. The base material consisted of slices of germanium 4 mils thick and $\frac{1}{4}$ inch square. The ohmic contact consisted of a circle of gold wire bonded into the germanium. Within this circle and on the same side of the germanium a regrown rectifying junction 120 mils in diameter, made by the indium alloying technique, was attached. Two germanium samples were used for the fabrication of these units, having resistivities of 4 ohm cm and 15 ohm cm, the latter being from a slice cut close to the n-type sample used in the experiments of Section B. Contacts were arranged to the gold and to the indium.

The opposite face of the germanium was masked with a piece of bakelite $\frac{1}{32}''$ thick, having a hole in the middle equal in size to the indium rectifying junction (Fig. 17). The bakelite was attached to the germanium surface with Ucilon lacquer. Care was taken to protect the edges of the unit with the same lacquer in order to insulate exposed parts of the gold spiral from contact with the electrolyte. The unit was set into the electrolyte as in Fig. 1, care being taken to see that no bubbles were left in the region of the exposed germanium surface. The tip of the salt bridge was set directly under the exposed region. With this geometry one had to be especially careful to see that current was not left on for too long in either direction in order to minimize the accumulation of
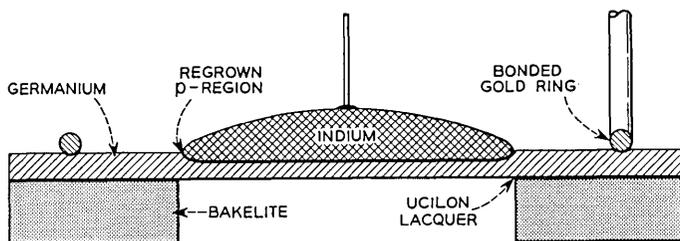
Fig. 17 — Diagram of thin-slice unit assembly.

gases or other reaction products in the confined space around the germanium.

3. PROCEDURE

3.1. *Light Calibration*

For calibration of the chopped light one can conveniently measure currents collected at the rectifying junction on a short-circuit basis (Fig. 18). For this purpose the unit may either be set into the electrolyte or left directly above it: in the latter case, however, the corrections for the change in reflection coefficient and for the removal of one reflecting surface are a little uncertain. For calibration of the settings of steady light which are used in the experiment, it is convenient to take a series of measurements of V^R, the voltage between the alloy junction and the base, as a function of I^R, the current flowing between the alloy junction and the base. By fitting the experimental points, with and without the steady light, on to the same rectifier curve, one may deduce the value of the photo current collected by the rectifying junction.

3.2. *Relation between floating potential and electrode potential, and determination of α by illumination*

The steady light is fixed at some value, determined as above. Measurements are made of electrode potential V^E and of the voltage V^R between the alloyed junction and the base, as functions of the electrolyte current. This is then repeated at other values of the steady light.

3.3. *Determination of α by current injection*

The above procedure is repeated, without steady light, at a series of values of current I^R across the alloyed rectfying junction.
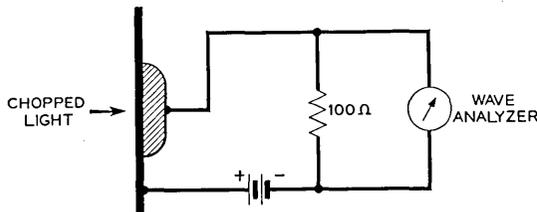
Fig. 18 — Electrical circuit for light calibration with thin-slice unit.

4. ANALYSIS OF THE EXPERIMENTS — RESULTS

In writing equation (2), we considered only minority carriers created at or very near the surface across which the electrolyte current is passing. We now wish to modify this equation to include minority carriers injected from the far surface, that is, across the rectifying junction. Since the floating potential at this junction measures the minority carrier density there rather than at the electrolyte surface, we should like also to be able to relate these two quantities. It may be shown (see Appendix 1) that the effect of recombination at the sides, in the body and at both surfaces may be described by writing

$$\left(\frac{p_1}{p}\right)^E = 1 + \frac{I_p^E + L + \gamma' I_p^R}{I_s} \tag{8}$$

$$\left(\frac{p_1}{p}\right)^R = 1 + \frac{\gamma''(I_p^E + L) + I_p^R}{I_s} \tag{9}$$

where γ' and γ'' are quantities slightly less than 1, analogous to the forward and reverse α's of a junction transistor. Since the slice of germanium is thin in comparison with a body diffusion length, and since the surface recombination velocities at the rectifying junction and at the electrolyte junction are fairly low, it will be sufficient to suppose that most of the generation and recombination takes place by diffusion to the sides. Under these conditions it is a sufficient approximation to write $\gamma' = \gamma'' = \gamma$. It is sufficient to suppose that

$$I_p^R = I^R$$

because it is known that almost all of the current crossing an indium regrown junction on n-type germanium is hole current, that is, the emitter efficiency is very close to 1.

We are now in a position to analyze experiments described in the preceeding paragraph.

(a) Set $I_p^E = 0$. Then, since

$$\left(\frac{p_1}{p}\right)^R = e^{\beta V_f}$$

and

$$V^A = V_f + I^R R_{\text{spr.}}$$

we have, from equation (9)

$$V^R = I^R R_{\text{spr.}} + \frac{1}{\beta} \ell n \left[1 + \frac{\gamma L + I^R}{I_s} \right] \tag{10}$$

Thus by taking a series of measurements of I^R and V^R we may determine the quantity γL for each light intensity.

(b) The difference between $(p_1/p)^E$ and $(p_1/p)^R$ will be relatively unimportant unless one of them is close to zero. Now, if the electrolyte current is near its saturation value, $(p_1/p)^E$ is effectively zero. But, however low the hole density at the lower surface may be, the hole density close to the rectifying junction can not be lower than is allowed by local hole generation. The lowest value of $(p_1/p)^R$ is determined by equations (8) and (9) to be

$$\left(\frac{p_1}{p}\right)^R = e^{\beta V_f{}^{\min}} = 1 - \gamma \tag{11}$$

Using equation (11) we may correct for the above effect and determine $(p_1/p)^E$ from V_f at all values of electrolyte current:

$$\left(\frac{p_1}{p}\right)^E = \frac{e^{\beta V_f} - e^{\beta V_f{}^{\min}}}{1 + e^{\beta V_f{}^{\min}}} \tag{12}$$

Using this equation one may take the experimental results and use them
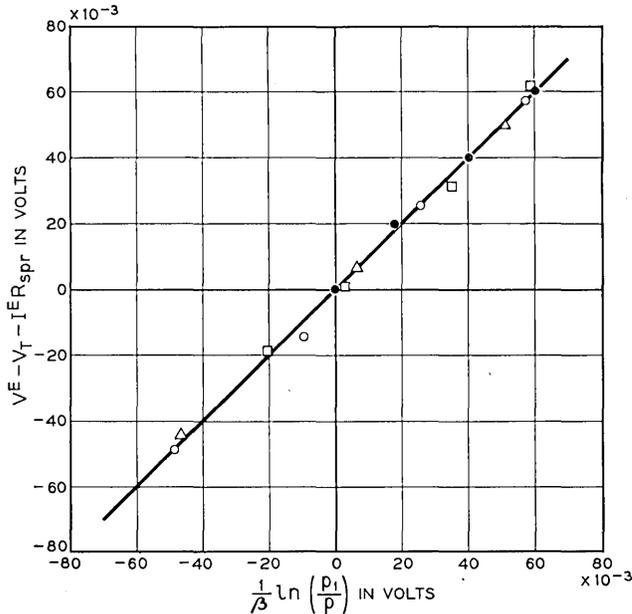
Fig. 19 — Plot of the quantity $(V^E - V_T - I^E R_{\text{spr.}})$ against $\beta^{-1} \ln (p_1/p)$.

to provide a direct verification of equation (1). This is shown for the 15 ohm cm sample in Fig. 19 in which we have chosen the Tafel over voltage parameter K_A in such a way that, by plotting the quantity against

$$\beta^{-1} \ell n \ (p_1/p)^E$$

against

$$[V^E - V_{\text{Tafel}} - I^E R_{\text{spr.}}]$$

the curves corresponding to different currents are superimposed on one another. It will be seen that the result is a straight line of slope unity. This may be regarded as the most direct experimental evidence for equation (1) available at the present time. The value of K_A used in the construction of this graph was 2.2 which is in reasonable agreement with that determined in Section B.

Two further observations are relevant. With the sample in the dark and no electrolyte current flowing, it was not possible to detect any significant departure of the floating potential from zero. This enables one to put an upper limit on the hole current which is flowing when the applied external current is zero as a result of some corrosive process. Since the smallest floating potential which could have been detected is of the order of 1 millivolt, one concludes that the maximum net hole current which could be flowing is $10^{-3} \beta I_s$, which is about $2\mu A/cm^2$. The other observation concerns the dependence of floating potential upon cathode current. After making due allowance for the potential
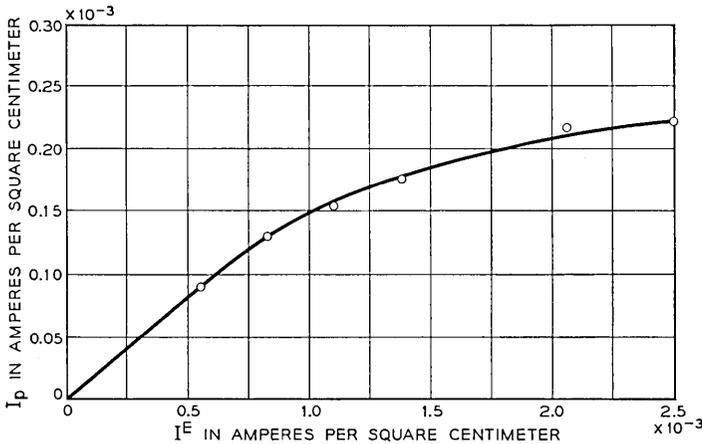
Fig. 20 — Injection of minority carriers on n-type germanium for cathodic currents.

drop across the spreading resistance, one may deduce the rate of in-
jection of holes as a result of the cathodic reaction at the surface. This
is shown in Fig. 20, in which the hole current as deduced from the meas-
urements and from equation (2) is plotted directly against cathode
current. It will be seen that the amount of injection is quite small. This
is in agreement with the general observation that the principal reaction
in the cathode direction is with electrons.

(c) From the measurements in various lights one may plot a curve
showing the limiting electrolyte current, at some suitable value of V^E,
as a function of the quantity γL deduced from the light calibration.
Assuming as usual that the quantity Δ (see equation 2) is a constant at
a constant electrode potential, one should have

$$I^E = -\frac{\alpha}{\gamma}(\gamma L) + \Delta - \alpha I_s \qquad (13)$$

Fig. 21 shows this curve plotted for a 4-ohm cm sample. From the slope
of this graph one has $\alpha/\gamma = 1.77$. A similar plot for the 15-ohm cm sample
gave $\alpha/\gamma = 1.90$. The values of γ as deduced from the minimum floating
potential (see above) were 0.83 and 0.95 respectively, thus giving $\alpha =$
1.48 on the 4 ohm cm sample and 1.80 on the 15 ohm cm sample. The
reason for this apparent dependence of α on resistivity is not known.

If in the same way we take the current-injection data, and plot I^E
against I^R at some fixed value of V^E, we should have:

$$I^E = -\alpha\gamma(I^R) + \Delta - \alpha I_s \qquad (14)$$

This plot is also shown in Fig. 21 for the 4-ohm cm sample. The slope
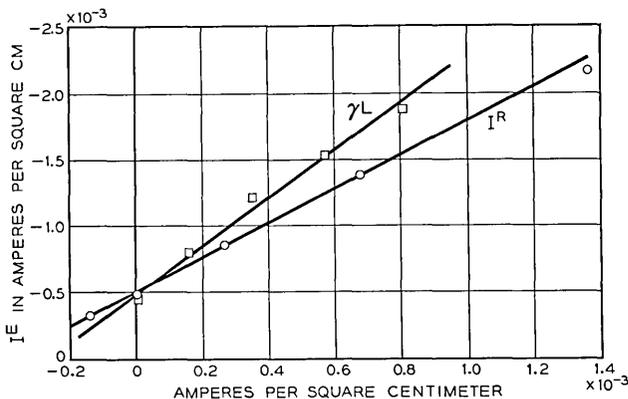
Fig. 21 — Plot of I^E as a function of γL or I^R.

of this curve should be $\alpha\gamma$. The experimental value is 1.30; this gives $\alpha = 1.56$, which checks well enough with the value obtained by illumination of the surface. It is also possible to deduce the quantity I_s for the thin slice geometry. However, as has been explained, the interpretation of this quantity is complicated by recombination at the edges of the sample and the flow is far from one-dimensional. In addition, the saturation in the thin slice experiment is usually poorer than that found on the large blocks. The reason for this is not known. It is possible that the sharpness of saturation may depend on the exact state of the surface; it is more difficut to prepare the thin slice samples in such a way that the surface is in a clean and reproducible condition. It is just possible that poorer saturation found with the thin-slice samples could be ascribed to leakage across the insulating lacquers.

D. The Region of Change in Sign of the Photo Voltage

In the preceding sections we have discussed the experimental results obtained for anodic currents with n-type material, and for cathodic currents with p-type material. Throughout these regions, the surface photo-voltage would have the constant value $1/\beta I_s$ if it were not for enhancement or reduction of the density of minority carriers near the surface. (Equation (4) and its analogue for p-type material.) Looking at Figs. 6 and 12, we may say that, if the experimental values found in the upper parts of the higher branch for each sample were corrected by multiplying by the appropriate ratio of minority carrier concentrations (and of course for imperfect saturation), the corrected photo-voltage would be constant throughout the regions mentioned, and would be approximately equal to those corresponding to the shoulders seen in the "dark" curve in Fig. 6 and in Fig. 12. Figs. 6 and 12 would then look very much like the curves of surface photo-voltage against contact potential given for the germanium-gas case by Brattain and Bardeen. We now turn our attention to the remaining sections of Figs. 6 and 12, in which the surface photo-voltage, quite irrespective of variations in minority carrier density, is changing with electrode potential.

We consider the ratio of the limiting anode and cathode values first, taking equations (1) and (6) as the basis for our discussion. Equation (1) says that the magnitude of anodic current flowing under given conditions is a function of the difference between V^E and the imref φ_{ps} for holes just inside the space-charge region, since φ_{ps} is related to the hole concentration at this point by the equation

$$\varphi_{ps} = \varphi_0 + \beta^{-1} \ln (p_1/p). \tag{15}$$

(φ_0 is the equilibrium hole imref.) If then the electrolyte current is held constant, and the hole imref changed (for example by shining light onto the sample), the change in electrode potential must be equal to the change in hole imref — a quantity which has the same sign for p as for n type, but a much smaller magnitude. On this basis we may derive the surface photo-voltage for both n and p type under conditions of anodic current (see Appendix 2):

$$\left(\frac{\partial V^E}{\partial L}\right)_I = \frac{1}{\beta I_{ps}{}^A} \text{ (n-type)} \qquad \left(\frac{\partial V^E}{\partial L}\right)_I = \frac{nb}{p}\frac{1}{\beta I_{ns}{}^A} \text{ (p-type)} \qquad (16)$$

for $p_1/p = 1$ (b = ratio of electron to hole mobility.) Here the suffixes n and p distinguish electron and hole saturation currents, and A specifies the recombination current, and therefore the surface recombination velocity, corresponding to *anodic* conditions.

Similarly, equation (6) says that the magnitude of cathodic current is a function of the difference between V^E and the imref φ_{ns} for electrons just inside the space-charge region. From an argument identical with that given above, we deduce the following values for the surface photo-voltage under cathodic conditions:

$$\left(\frac{\partial V^E}{\partial L}\right)_I = -\frac{p}{nb}\frac{1}{\beta I_{ps}{}^C} \text{ (n-type)} \qquad \left(\frac{\partial V^E}{\partial L}\right)_I = -\frac{1}{\beta I_{ns}{}^C} \text{ (p-type)} \qquad (17)$$

for $n_1/n = 1$.

Using these relations, we may compare the minority carrier saturation currents in the anode and cathode directions for the two samples:

$$\frac{I_{ns}{}^C}{I_{ns}{}^A} = -\frac{(\partial V^E/\partial L) \text{ an}}{(\partial V^E/\partial L) \text{ cath}}\bigg/\left(\frac{nb}{p}\right) = \frac{100}{40} = 2.5$$

$$\frac{I_{ps}{}^C}{I_{ps}{}^A} = -\frac{(\partial V^E/\partial L) \text{ cath}}{(\partial V^E/\partial L) \text{ an}}\bigg/\left(\frac{p}{nb}\right) = \frac{180}{100} = 1.8$$

$$(18)$$

Within the accuracy of the experiment, we conclude that the ratio is about 2 in each case. Since surface recombination velocity at the electrolyte surface accounts for only part of the total minority carrier recombination current, the ratio of the surface recombination velocities in the cathode and anode conditions must be somewhat larger than this. For the balance of this section, however, we shall ignore the difference between the anodic and cathodic values of I_s .

In the intermediate region, the simplest interpretation is that the electrode reaction is neither wholly anodic nor wholly cathodic, irrespective of whether the *net* electrolyte current is anodic or cathodic. We sup-

pose that the anodic and cathodic processes are still described by equation (1) and (6), but that the net current is now the sum of the currents appearing in those two equations. Dropping the IR terms from equations (1) and (6), and rewriting in terms of currents instead of voltages, we have:

$$I_A{}^E = I_{0A}[\exp{[- (\beta/K_A)\{V^E - V_{0A} - (\varphi_{ps} - \varphi_0)\}]} - 1]$$

$$I_C{}^E = I_{0C}[\exp{[+ (\beta/K_C)\{V^E - V_{0C} - (\varphi_{ns} - \varphi_0)\}]} - 1] \qquad (19)$$

$$I^E = I_C{}^E - I_A{}^E$$

With $\varphi_{ps} = \varphi_{ns} = \varphi_0$, equations (19) have the form of two imperfect rectifiers in shunt, arranged in the opposite sense one to another; a large voltage applied in either direction will cause a large current to flow through one, while only the comparatively small "saturation current" (I_{0A} or I_{0C}) flows through the other. (This is why we are able to ignore cathodic processes when there is a large current in the anodic sense and vice versa.) Using the method given in Appendix 2, we may show for n-type semiconductor in the limit $p_1/p = 1$ and $n_1/n = 1$, which is certainly appropriate to the measurements made at small currents and with no steady light,

$$\left(\frac{\partial V^E}{\partial L}\right)_{I^E} = \frac{1}{\beta I_s}\frac{1 - (X/X_0)}{1 + (nb/p)(X/X_0)}$$

where $\qquad (20)$

$$X = \exp\left[\beta\left(\frac{1}{K_A} + \frac{1}{K_C}\right)V^E\right], \qquad \text{and}$$

X_0, which is the value of X for which the surface photo-effect vanishes, is given by:

$$X_0 = \frac{bn}{p}\frac{I_{0A}}{I_{0C}} e^{\beta((V_{0A}/K_A)+(V_{0C}/K_C))} \qquad (21)$$

The expressions for p-type are analogous.

This may be tested by plotting the quantity

$$Q = \ell n\left\{\left[1 - \frac{(\partial V/\partial L)}{(\partial V/\partial L)_{\text{cath.}}}\right]\bigg/\left[1 - \frac{(\partial V/\partial L)}{(\partial V/\partial L)_{\text{an}}}\right]\right\}$$

against V^E (Fig. 22). This gives a straight line for both samples; from the slope one deduces $1/K_A + 1/K_C = 1.0$. The value of K_A deduced from the measurements at larger anodic currents was about 2; this suggests that K_C is about 2 also. (Note that this is not necessarily the same
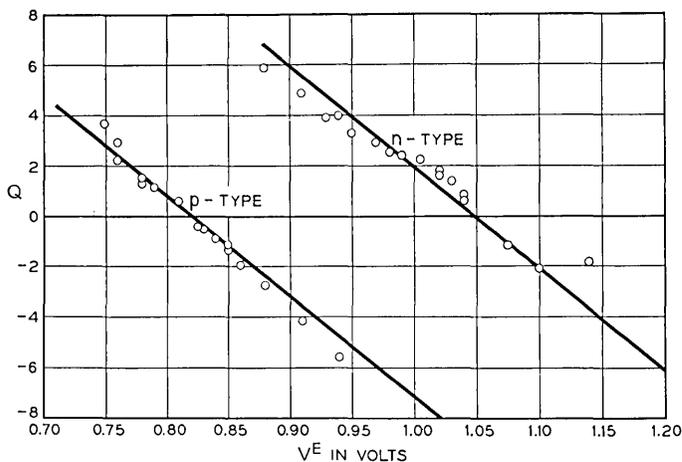
Fig. 22 — Plot of the quantity Q against electrode potential for p and n type germanium.

cathodic process as that represented by the Tafel curve in Fig. 15, since this is below, the other above, the region of voltage instability.)

Equation (21), and its analogue for p-type material, depends on resistivity only through the factor (bn/p). To test this, we make use of the difference between the values of V^E at which the surface photo-voltage vanishes for the n and for the p-type sample. Using the value of $(1/K_A + 1/K_C)$ deduced above, we may construct the following table:

Sample	n-type	p-type
V^E for $I^E = 0$	0.83 volts	0.84 volts
V^E for which $(\partial V/\partial L) = 0$	1.05	0.82
$\dfrac{K_A K_C}{K_A + K_C} \dfrac{1}{\beta} \ell n \dfrac{(\partial V/\partial L)_{\text{cath}}}{(\partial V/\partial L)_{\text{an}}}$	−0.12	0.12
Sum of second and third lines	0.93	0.94

The voltages in the last line of the table are, within the experimental error, the same for the two samples, thereby checking (as one would expect) that the purely electrochemical features of the electrode processes near zero current, as opposed to those which depend on the hole or electron concentrations, are independent of the type of germanium used.

It is not possible to deduce the separate values of the quantities V_{0A}, V_{0C}, I_{0A}, and I_{0C}. In principle, the equilibrium potentials could be

deduced from thermodynamical reasoning from heats of reaction, but we are ignorant as to the exact nature of the reactions taking place. We can however draw one rather interesting conclusion from the fact that the zero in photo-voltage for a sample having $nb/p = 1$ occurs at a potential about 0.1 volts higher than that corresponding to zero current. This one may understand as follows. When the net current through the surface is zero, the anodic and cathodic currents $I_A{}^E$ and $I_C{}^E$ are equal in magnitude and opposite in sign. (We know from the work of Section C that each is of the order of $1\mu A/cm^2$ or smaller.) Since in this state the surface is on the anodic or p side of the zero in photo voltage, it follows that, if a small current is allowed to cross the surface, almost all of it will be represented by change in $I_A{}^E$ and very little by a change in $I_C{}^E$. We may in fact show from equations (19) and (21) that, if δV represents the difference between electrode potential corresponding to zero current, and that at which the photo-voltage vanishes for a sample having $p/nb = 1$, then the ratio of cathodic (electron) to anodic (hole) current for a small displacement from the zero-current condition is given by:

$$\frac{K_c}{K_A} e^{\beta((1/K_A)+(1/K_C))(\delta V)} \tag{22}$$

From the experimental values, this ratio is about 0.03. This, of course, is true whether the net current into the germanium from the solution is positive or negative. If we may identify the anodic process with oxidation of germanium, this result shows that, at sufficiently small currents, the predominant process is either oxidation of germanium or reduction of germanium oxide, depending on the sign of the external current, and that it is only at higher cathodic currents that some other process, represented by $I_C{}^E$, takes over. Whether or not the alternative process is plating out of hydrogen (which should take place somewhere near this voltage) we do not know.

It should be mentioned here that the voltage difference δV does not vary much with electrolyte, although the separate voltages may change considerably. One reason for this may be that V_{0A} and V_{0C} are shifted by equal amounts when the pH of the solution is changed. One may tentatively conclude that the surface of a sample of germanium in contact with an aqueous electrolyte, with no net current crossing the surface, is p-type in character whether the measured electrode potential is high or low.

E. Concluding Remarks

(1) The primary reaction at a germanium anode is with holes rather than with electrons in the semiconductor. The evidence for this is that

the measured electrode potential in the anode direction depends on the concentration of holes near the surface in the way to be expected from the purely thermodynamical reasoning given in Appendix 2.

2. Observed values for the current gain at an anode composed of n-type germanium lie between 1.4 and 1.8. One must suppose that a product of the primary reaction is itself unstable, and decomposes in such a way as to allow electrons to go into the conduction band of the semiconductor. We believe that current multiplication of some such sort as this plays a very important role in other electrolytic interface systems.

3. The initial primary reaction at a germanium cathode is with electrons rather than with holes in the semiconductor. Sometimes this state of affairs persists; sometimes the reaction seems to be replaced, in a time of the order of seconds or minutes, by one involving holes. The current gain is of the order of unity.

4. The surface recombination velocity at the surface of an n-type or p-type germanium sample of the resistivities used, in contact with an electrolyte, is not greater than a few hundred cm/sec.; and it is less when the surface is passing anodic current than cathodic. At the same time we do not exclude the possibility that, at large cathodic current densities, or after a considerable interval of time, the surface recombination velocity, depending on the way in which it is defined, is large. Some peculiar effects have been observed by us in a different experiment from that described here. An attempt was made to measure minority carrier lifetime by exposing a rod of n-type germanium to a cathodic current and studying the decay of photo-conductivity (the Haynes-Hornbeck[15] experiment). The decay curve showed a very sharp initial decline, followed by a horizontal portion, and then a much slower disappearance of remaining carriers. The reason for this is not understood. It is perhaps worth noting here that some time was wasted in searching for a reliable way to determine the surface recombination velocity while the surface was exposed to an electrolyte by such methods as measuring diffusion lengths in filaments. In interpreting such experiments, one must not forget the extra complications arising from current crossing the surface.

5. There are some similarities between the properties of semiconductor-electrolyte interfaces and those of semiconductor-metal junctions. In both cases it is possible to obtain rectification on both n and p-type germanium, with that on n-type being sharper. In both cases there can be current gain. The geometry of a point-contact rectifier is very difficult to analyze quantitatively. We suggest that the techniques discussed in this paper might however prove profitable in understanding the type

of plated metal-semiconductor contact recently described by Schwartz and Walsh[15] and others.[16]

6. The rectifying properties of a semiconductor-electrolyte interface may also be described as follows. When n-type germanium is passing anodic current, it is of such a sign as to attract negative ions to the surface; in consequence, the sign of the space-charge region in the surface of the semiconductor must be positive. One might therefore expect that the surface would be p-type. Since the bulk and the surface of the semiconductor now form a p-n junction, across which current is passing in the reverse direction, the current eventually saturates. This description, however, obscures the main point: the current across a junction only saturates when the current from *both* sides is limited by diffusion, and it is again necessary to conclude that electrons cannot easily pass from the anions in the solution into the semiconductor.

7. The relation found experimentally between surface photo-effect and electrode potential is strikingly similar to that reported by Brattain and Bardeen between surface photo-effect and contact potential in the germanium-gas system. Nevertheless, the interpretation we have given appears to be quite different. Let us therefore see how the treatment of the contact potential experiments might apply to the present experiments. In the theoretical discussion of Brattain and Bardeen, the point of departure is that the total surface *charge-density* due to the space-charge and surface states associated with the germanium and to ions adsorbed on the surface, is zero. When the surface is suddenly illuminated, the last contribution does not have time to change: the surface photo-effect may then be calculated on certain assumptions. Now it would seem to be perfectly possible to treat the germanium-electrolyte system in the same way. Any such model must agree with the thermodynamical treatment in the extremes of electrode or contact potential: it is only in the intermediate region that there can be discrepancies. In the paper of Brattain and Bardeen,[6] where space-charge is neglected, the extent of the intermediate region depends on the number of surface states; in a treatment of Garrett and Brattain,[7] where surface states are neglected, the extent of the intermediate region is fixed. We have evaded the issue altogether in this paper, by employing semi-empirical expressions for the over-voltage, equations (1) and (6). We have not attempted to describe in detail the problem of the relation between "activation" and current crossing the surface, from which the expressions for over-voltage can be derived, this being the problem in the electrolyte system analogous to that of the relation between the measured contact

potential and the surface potential (the electrostatic potential at the semiconductor surface) in the germanium-gas case. One curious fact may however be pointed out. The theoretical treatment of Garrett and Brattain gives the relation between surface photo-effect and surface potential in the same form as equation (20) of this paper, except that the V^E is replaced by Ψ, the difference between the electrostatic potential at the surface and in the interior, and the coefficient

$$\left(\frac{1}{K_A} + \frac{1}{K_c}\right)$$

is replaced by unity. Now we found the experimental value of

$$\left(\frac{1}{K_A} + \frac{1}{K_c}\right)$$

to be 1.0; it is therefore tempting to apply the theory of Garrett and Brattain to the present case. To do this would be to claim that V^E and Ψ are identical, except for an additive constant, in the region of low current density: that is, all the measured changes in electrode potential occur across the space-charge region, and effectively none across the space-charge region in the electrolyte. Whether such a conclusion is justified must be left for further experiments to decide. In any case it has to be admitted that, since the interpretation of these experiments may be so concisely accomplished by thermodynamical arguments, the role of surface states remains obscure. However, the main deduction of this work — the importance of holes in anodic processes and of electrons in cathodic processes — may furnish a clue towards the understanding of more complicated effects of chemical treatments upon semiconductor surface properties.

APPENDIX 1

ANALYSIS OF THE THIN SLICE EXPERIMENTS

We start with some remarks about surface recombination velocity. Consider a plane surface bounding n-type germanium; let x be the distance between a point in the germanium and the surface. Let $x = X$ define a plane just sufficiently far into the material that the field E is comparatively small, i.e. such that, in any flow of minority carriers, diffusion predominates over drift. We shall refer to a point in this plane as being "just inside the space-charge region". Experiment shows that the recombination current to the surface depends on the concentration p_1^X of minority carriers at X. The surface recombination velocity v is defined[17] so that

$$(I_p)^{\text{rec.}} = ev(p - p_1^X) \tag{A1}$$

where p is the equilibrium concentration, and $(I_p)^{\text{rec.}}$ is the hole current flowing away from the surface as the net result of surface thermal generation and surface recombination.

Suppose next that current is crossing the surface, in such a way as to tend to change the minority carrier density in the semiconductor. This state of affairs may be described by replacing (A1) by the equation:

$$I_p = I_p' + ev(p - p_1^X) \tag{A2}$$

where I_p' may be considered to be "net hole current flowing into the semiconductor". Notice that, even if one can measure I_p with and without current crossing the surface, equation (A2) does not suffice to determine I_p' experimentally, since we have said nothing as to whether v is or is not also changed by switching on the current. (If, however, we assume that I_p' is uniquely determined by the current crossing the surface, we may vary p_1^X and so determine v. This is what is done in the preceding paper.)

We turn now to the problem of illumination of the surface. Visible light is absorbed in a thickness comparable with that of the surface space-charge region. Let us suppose that the hole current very close to $x = 0$ is still given by equation (A2): what then is the hole current at $x = X$? First notice that X, although large in comparison with atomic dimensions, is (in germanium) very small in comparison with a body diffusion length. Therefore it is sufficient to assume that there is no recombination between 0 and X. Throughout this region, p_1 may vary with distance in a complicated way, and may even be greater than n_1

near the surface; but the continuity equation:[18]

$$\frac{\partial p_1}{\partial t} = \frac{1}{e} \operatorname{div} I_p - f(x) \tag{A3}$$

is always satisfied. Here $f(x)$ stands for the rate of creation of holes per unit volume by photo-excitation at $x = X$. $f(x)$ satisfies the condition

$$N = \int_0^\infty f(x)\, dx \sim \int_0^X f(x)\, dx \tag{A4}$$

where N is the rate of pair excitation per unit area of surface. In the steady state $\partial p_1/\partial t = 0$, and for our one-dimensional geometry we have

$$I_p = e \int_0^X f(x)\, dx + A \tag{A5}$$

But we suppose that, at $x = 0$, I_p is given by equation (A2): therefore

$$(I_p)_x = eN + I_p' + ev(p - p_1^{\,X}) \tag{A6}$$

But X has been chosen in such a way that E is negligible in the flow of minority carriers: thus we may write:

$$\left(\frac{dp_1}{dx}\right)_x = -\frac{(I_p)_x}{eD} = -\frac{L + I_p'}{eD} + \frac{v}{D}\,(p_1^{\,X} - p) \tag{A7}$$

We now proceed to use this as the appropriate boundary condition in the solution of the minority carrier flow problem for the thin slice geometry. In this problem we suppose the germanium to be in the form of a thin sheet of thickness t, extending indefinitely in all directions. The electrolyte and alloyed contacts are in the form of two circular areas, of radius r_0, opposite one another, one on each side of the sheet. Recombination and generation of minority carriers can in principle take place (i) in the body; (ii) at the germanium surface lying further than r_0 from the axis; (iii) at the electrolyte interface; and (iv) at the alloyed interface. If it were not for (ii) the problem would be one-dimensional, and quite easy to solve. The whole problem is of course identical with that of a junction alloyed transistor.

Consider first diffusion to the sides, with a recombination velocity v_0 for the neighboring germanium. Assume as a rough approximation that the hole density is constant across the slice, i.e., is a function only of r, the distance from the axis. Elementary reasoning then leads to the equation:

$$r^2 \frac{d^2\Delta p}{dr^2} + r \frac{d\Delta p}{dr} - \frac{2v_0}{tD}\,r^2\Delta p = 0 \tag{A8}$$

where $\Delta p = p_1 - p$. The solution of this having the property $\Delta p \to 0$ as $r \to \infty$ is the Hankel function of order zero:

$$\Delta p = C H_0^{(1)} \left(i r_0 \sqrt{2v_0/tD} \right) \tag{A9}$$

Let us define a quantity v^*, such that the total flow of minority carriers across the cylinder of radius r_0 is given by:

$$(2\pi r_0 t) e v^* (\Delta p)_{r_0} \tag{A10}$$

From equation (A9) one finds

$$v^* = \frac{D}{L_s} \left[- H_1^{(1)} \left(\frac{i r_0}{L_s} \right) \bigg/ H_0^{(1)} \left(\frac{i r_0}{L_s} \right) \right] \tag{A11}$$

where L_s is the "sideways diffusion length":

$$L_s = \sqrt{tD/2v_0} \tag{A12}$$

In a practical case L_s is rather smaller than r_0, and the ratio of the Hankel functions in equation (A11) is not very different from unity. The purpose of the foregoing analysis is to show that, at least in the approximation used, recombination at the outlying free germanium surface may be treated by regarding the surface of the cylinder of radius r_0 as the recombining surface, with a recombination velocity given by equation (A11).

We now use this concept to apply to the case in which the hole density is allowed to vary across the slice, and, as we are now mainly interested in condition for $r < r_0$, we suppose that now the radial variation in p may be neglected. From the diffusion and continuity equations one has:

$$D \frac{d^2 \Delta p}{dx^2} = \left(\frac{2v^*}{r_0} + \frac{1}{\tau} \right) \Delta p \tag{A13}$$

The solution of this is

$$\Delta p = A e^{x/\lambda} + B e^{-x/\lambda} \tag{A14}$$

where

$$\lambda^2 = D \bigg/ \left(\frac{1}{\tau} + \frac{2v^*}{r_0} \right) \tag{A15}$$

A and B are determined by the boundary conditions at* the electrolyte surface ($x = 0$) and at the alloyed surface ($x = t$). Let us suppose that the recombination velocities are v_E and v_A, the latter being, pre-

---

* I.e., "just inside".

sumably, small: and suppose that the hole current across the electrolytic surface (in the sense defined above) is $I_p{}^E$, and that across the regrown surface $I_p{}^R$: and suppose finally that the electrolyte surface is illuminated, with L/e hole-electron pairs created per second. $I_p{}^E$, $I_p{}^R$ and L are expressed on a unit-area basis. From the boundary conditions (as in equation A7) we have:

$$-eD \left.\frac{dp_1}{dx}\right)_0 = -\frac{eD}{\lambda} (A - B)$$

$$= (I_p{}^E + L) - ev_E(p_1{}^0 - p)$$

$$-eD \left.\frac{dp_1}{dx}\right)_t = -\frac{eD}{\lambda} (Ae^{t/\lambda} - Be^{-t/\lambda})$$

$$= I_p{}^R + ev_A(p_1{}^t - p)$$

$$(A16)$$

We are interested in $p_1{}^t$ and $p_1{}^0$. Expanding the exponentials, and performing a certain amount of algebra, one finds:

$$\frac{p_1{}^0}{p_0} = 1 + \frac{\gamma' I_p{}^R + (I_p{}^E + L)}{I_s}$$

$$\frac{p_1{}^t}{p_0} = 1 + \frac{\gamma''(I_p{}^E + L) + I_p{}^R}{I_s}$$

$$(A17)$$

where

$$\gamma' \sim 1 - \frac{t^2}{2D\tau} - \left(v^* \frac{t}{r_0} + v_R\right)\frac{t}{D}$$

$$\gamma'' \sim 1 - \frac{t^2}{2D\tau} - \left(v^* \frac{t}{r_0} + v_E\right)\frac{t}{D}$$

$$I_s = ep\left(\frac{2v^*t}{r_0} + \frac{t}{\tau} + v_R + v_E\right)(1 + \Delta)$$

$$(A18)$$

where Δ is small of the order of v^*t/D or a comparable quantity. Equations (17) are identical with equations (8) and (9) in the text.

APPENDIX 2

THERMODYNAMICAL DISCUSSION OF A SEMICONDUCTOR ELECTRODE

First we show how, when the germanium-electrolyte interface is permeable only to holes, the equilibrium electrode potential includes the

electrochemical potential for holes. We follow the notation of Guggen-heim[19] and allow μ to stand for a chemical potential of an uncharged species, and for an electrochemical potential of a charged species, both quantities being defined by the equation

$$\mu_i = (\partial G/\partial n_i)_{T,P,n_j} \tag{A19}$$

For holes and electrons in germanium

$$\mu_p{}^{Ge} = \varphi_p = \psi + \frac{1}{\beta} \ln \frac{p}{n_i}$$

$$\mu_n{}^{Ge} = -\varphi_n = -\psi + \frac{1}{\beta} \ln \frac{n}{n_i}$$

For germanium under equilibrium conditions, and for a metal under any practically realizable condition,

$$\mu_p + \mu_n = 0 \tag{A20}$$

Now consider the cell:

Metal$_1'$ | Dark Ge | Illuminated Ge | Solution | Metal$_2$ | Metal$_1''$

Suppose that holes are required for the reaction at the germanium solu-tion interface; it is immaterial which is required at the (metal$_2$ | solution) interface, since equation (A20) is satisfied in metal$_2$. Let us suppose that the cell reaction is:

$$A - p_{Ill\ Ge} = B + n_2 \tag{A21}$$

where A and B denote any chemical or ionic species. We have written the cell reaction in a rather unusual way, to emphasize that, at the same time as one gram-equivalent of A is converted into B, one Faraday of holes is removed from the germanium surface, and one Faraday of elec-trons removed from metal$_2$. Let us contemplate the process of taking one electronic charge across the cell, completing the circuit by connecting the cell terminals via a battery of e.m.f. equal to that of the cell. Since this process may be carried out reversibly, and since we have included

electrons and holes as thermodynamic components,

$$\Delta G = 0 \tag{A22}$$

Let us now write down the available thermodynamic relations between the μ's. To avoid conversion factors, we shall suppose them to be specified in electron volts. We have:

$$\mu^A - \mu^B - \mu_{ps}{}^{Ge} - \mu_{n2} = 0$$

$$\begin{aligned} \mu_{n2} &= \mu_{n1''} \\ \mu_{p0}{}^{Ge} &= \mu_{p1'} \\ \mu_{p1'} &= -\mu_{n1'} \end{aligned} \tag{A23}$$

where $\mu_{ps}{}^{Ge}$ and $\mu_{p0}{}^{Ge}$ refer respectively to the surface (illuminated) and equilibrium (dark) parts of the germanium. Then the measured voltage between the right and left hand sides of the cell is

$$\begin{aligned} V &= \mu_{n1'} - \mu_{n1''} \\ &= \mu_B - \mu_A + (\mu_{ps}{}^{Ge} - \mu_{p0}{}^{Ge}) \end{aligned} \tag{A24}$$

Thus, so long as the potentials of the reacting components in the solution remain constant,

$$V = V^0 + \frac{1}{\beta} \, \ell n \, \frac{p_1}{p} \tag{A25}$$

The same argument may be repeated for the case that the germanium-electrolyte interface is permeable only to electrons, and gives

$$V = V^{0'} - \frac{1}{\beta} \, \ell n \, \frac{n_1}{n} \tag{A26}$$

We now evaluate the magnitude of the photo-voltage for these two extreme conditions:

$$\begin{aligned} \left(\frac{\partial V}{\partial L}\right)_I &= \frac{\partial}{\partial L}\bigg)_I (\varphi_{ps} - \varphi_0) \qquad \text{anode} \\ \left(\frac{\partial V}{\partial L}\right)_I &= \frac{\partial}{\partial L}\bigg)_I (\varphi_{ns} - \varphi_0) \qquad \text{cathode} \end{aligned} \tag{A27}$$

If the material is n-type, and if the mobility of electrons and holes were equal, it may be shown[7] that

$$\left.\frac{\partial}{\partial L}\right)_I (\varphi_{ps} - \varphi_0) = \frac{1}{\beta}\frac{1}{I_s + L + I_p} = \frac{1}{\beta I_s}\frac{p}{p_1}$$

$$\left.\frac{\partial}{\partial L}\right)_I (\varphi_{ns} - \varphi_0) = -\left(\frac{p}{n}\right)\frac{1}{\beta}\frac{1}{I_s + (p/n)(L + I_p)} \qquad (A28)$$

$$= -\frac{p}{n}\frac{1}{\beta I_s}\frac{n}{n_1}$$

with the notation used in the text. Because the mobilities are not equal, there is an additional term,[20] too small to affect the upper equation, but of the same magnitude as the lower:

$$\frac{d\delta V_i}{dL} = +\left(\frac{p}{n}\right)\frac{1}{\beta}\left(\frac{b - 1}{b}\right)\frac{1}{I_s + \frac{p}{n}\frac{b + 1}{b}(L + I_p)} \qquad (A29)$$

Allowing for this, and approximating equations (A28) and (A29) for the case that L and I_p are small in comparison with I_s,

$$\left(\frac{\partial V}{\partial L}\right)_I = \frac{1}{\beta I_s} \qquad \text{anode}$$

$$\left(\frac{\partial V}{\partial L}\right)_I = -\frac{p}{nb}\frac{1}{\beta I_s} \qquad \text{cathode} \qquad (A30)$$

In a similar way, it may be shown for p-type material that

$$\left(\frac{\partial V}{\partial L}\right)_I = \frac{nb}{p}\frac{1}{\beta I_s} \qquad \text{anode}$$

$$\left(\frac{\partial V}{\partial L}\right)_I = -\frac{1}{\beta I_s} \qquad \text{cathode} \qquad (A31)$$

under the same conditions. (I_s now means the *electron* saturation current.) Notice from equation (28) that the cathode photo-effect on n-type (and the anode photo-effect on p-type) is relatively insensitive to the magnitude of L and I_p. This is well shown in Fig. 6.

BIBLIOGRAPHY

1. Brattain, W. H., and Garrett, C. G. B., Phys. Rev., **94**, p. 750, 1954.
2. Mott, N. F., Proc. Roy. Soc., A **171**, p. 27, 1939.
3. Schottky, W., Zeit. Phys., **118**, p. 539, 1942.
4. Shockley, W., and Pearson, G. L., Phys. Rev., **74**, p. 232, 1948.
5. Bardeen, J., Phys. Rev., **71**, p. 717, 1947.
6. Brattain, W. H., and Bardeen, J., B. S. T. J., **32**, p. 1, 1953.
7. Garrett, C. G. B., and Brattain, W. H., To be published.
8. Brattain, W. H., *Semiconducting Materials*, Butterworth, 1951, p. 37.
9. Becquerel, Edmond, Comptes Rendus, **9**, p. 561, 1839.

10. Glasstone, *Introduction to Electrochemistry*, Van Nostrand, 1942, Chapter XIII.
11. IBID, Chapter VII.
12. Shockley, W., B. S. T. J., **28,** p. 435, 1949.
13. Bardeen, J., B. S. T. J., **29,** p. 469, 1950.
14. van Roosbroeck, W., J. Appl. Phys. in process of being published.
15. Haynes, J. R., and Hornbeck, J. A., Phys. Rev., **90,** p. 152, 1953.
16. Schwartz, R. F., and Walsh, J. F., P.I.R.E., **41,** p. 1715, 1953.
17. Shockley, W., *Electrons and Holes in Semiconductors*, Van Nostrand, 1950, p. 321.
18. IBID, p. 320.
19. Guggenheim, E. A., *Thermodynamics*, Elseviet, 1949, Chapter X.
20. van Roosbroeck, W., B. S. T. J., **29,** p. 560, 1950.

Boolean Matrices and the Design of Combinational Relay Switching Circuits

By FRANZ E. HOHN and L. ROBERT SCHISSLER

(Manuscript received August 30, 1954)

The p-terminal generalization of a two-terminal switching function is a matrix of switching functions representing the conditions under which the terminals are interconnected. The properties of these "switching matrices" are studied, and examples are given to show how they may be employed effectively in the design of switching circuits. Some basic problems are outlined and a bibliography is attached.

CONTENTS

1. INTRODUCTION

Matrices over a Boolean algebra, or simply *Boolean matrices*, are rectangular arrays of elements from a Boolean algebra. These arrays are subject to appropriate rules of operation, some of which are analogous to the rules of operation for ordinary matrices, whereas others reflect the

Boolean character of the elements. The purpose of this paper is to present those properties of Boolean matrices which have application to the design of combination relay logic circuits, and to develop fundamental aspects of this application.

In this paper, we shall assume a knowledge of the elementary aspects of Boolean algebra and of its application to the design of switching circuits.[1,2] We shall use the following notations for the Boolean operations:

$+$ denotes the Boolean sum or union,

$\cdot$ denotes the Boolean product or intersection,

$\leqq$ denotes inclusion.

In order to be able to operate most naturally with Boolean matrices, we use the system in which the parallel connection of contacts x and y is represented by $x + y$ while their series connection is represented by $x \cdot y$ or just xy. We use 0 to denote an open circuit or contact and 1 to denote a closed circuit or contact. The Boolean algebra from which the elements of our Boolean matrices are selected is the set "S" of 2^{2^n} Boolean or switching functions of n variables x_1, x_2, $\cdots$, x_n.

2. MATRICES ASSOCIATED WITH COMBINATIONAL CIRCUITS

We shall be concerned with the analysis and synthesis of combinational relay circuits, that is, circuits which may be represented symbolically as in Fig. 1. Here there are indicated n coils x_1, x_2, $\cdots$, x_n, which determine respectively the conditions x_1, x_2, $\cdots$, x_n of contacts in the box. We call x_1, x_2, $\cdots$, x_n the *inputs* or *input variables* of the circuit. The outputs of the circuit are the interconnections between the terminals 1, 2 $\cdots$, p, which are established as a result of energization of certain of the coils. It is assumed that the contacts of the circuit are all in the box and that these contacts are operated solely by the coils
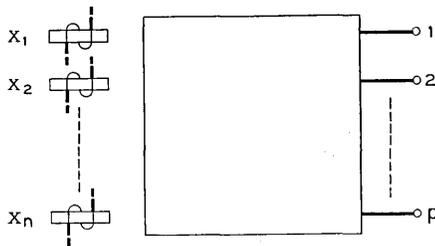
Fig. 1

X_1, X_2, $\cdots$, X_n, which are controlled entirely from outside the box. The design of such circuits is discussed in Reference 3, Chapter 6.

In a combinational circuit, after a brief operate-time, the state f_{ij} of the connection between the terminals i and j depends only on the combination of values assumed by the input variables x_1, x_2, $\cdots$, x_n, and hence may be represented as a Boolean function of these variables:

$$f_{ij} = f_{ij}(x_1, x_2, \cdots, x_n).$$

(Since a terminal i is always connected to itself, we define $f_{ii} = 1$ for each i.) The p^2 functions so obtained may be used as the elements of a $p \times p$ symmetric Boolean matrix which we call the output matrix "F" of the circuit:

$$F = [f_{ij}].$$

Thus, for example, the output matrix of a simple, three-terminal circuit is illustrated in Fig. 2. (In Fig. 2 and in succeeding figures, small rings are used to denote the p terminals of a network, and black dots are used to denote non-terminal nodes which simply serve as connecting points.)

Two p-terminal combinational circuits with the same output matrix are called *equivalent*. Equivalence is denoted by the symbol "$\sim$" in this paper.

With a given circuit, we can associate a second type of Boolean matrix in the following manner. First we select and number certain nodes in the circuit, using the numbers $p + 1$, $p + 2$, $\cdots$, $p + k$. These we call *non-terminal nodes* to distinguish them from the *terminal nodes* $1, 2, \cdots, p$ of the circuit. The non-terminal nodes are so chosen that between any two of the $p + k$ nodes of the circuit there appears at most a single contact or a group of single contacts in parallel. Moreover, we assume that every
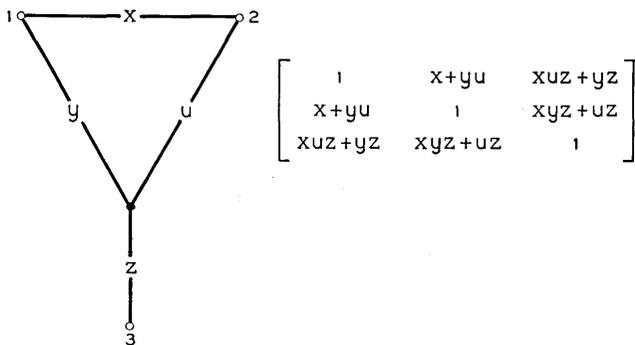
Fig. 2

contact of the network is included in the connection between some pair of nodes.

Let p_{ij} represent the "connection" between nodes i and j. This symbol has the value 0 if there is no connection at all and 1 if there is a short circuit, but otherwise it is the symbol denoting a single contact or is a sum of such symbols. The matrix "P" of order $p + k$:

$$P = [p_{ij}]$$

is then called a *primitive connection matrix* of the circuit. For example, if in Fig. 2 we select a single non-terminal node in the obvious way and number it "4", we obtain the primitive connection matrix

$$P = \begin{bmatrix} 1 & x & 0 & y \\ x & 1 & 0 & u \\ 0 & 0 & 1 & z \\ y & u & z & 1 \end{bmatrix}.$$

A third type of matrix which we often associate with a circuit falls, in a sense, between the primitive connection matrix and the output matrix. We call it just a *connection matrix*, "C". (This term was originated by Warren Semon at the Harvard Computation Laboratory.) In such a matrix, the entries are switching functions of two-terminal circuits connecting the nodes, both terminal and non-terminal, of the circuit, but the number of non-terminal nodes selected need not be large enough to lead to a primitive connection matrix. However, it is assumed that enough non-terminal nodes are selected so that all the contacts of the network are accounted for in the resulting two-terminal circuits. The situation is illustrated in Fig. 3. For certain purposes, a connection matrix such as this is as useful as a primitive one, or more so.

The output matrix of a circuit may also be regarded as a connection matrix if that appears desirable, but its primary importance lies in the fact that it is *the generalization of the switching function of a two-terminal circuit*. (Note that the output matrix of a two-terminal circuit is simply

$$F = \begin{bmatrix} 1 & f \\ f & 1 \end{bmatrix}$$

where f is the switching function of the circuit, and thus tells us nothing more nor less than the switching function f itself.)

The fundamental problem of the *analysis* of a combinational circuit involves writing a connection matrix corresponding to the circuit, there-

after deducing the corresponding output matrix or other relevant information. This presents relatively few difficulties. The problem of *synthesis*, on the other hand, is much more difficult, for it involves translating given operate-requirements into the form of an output or connection matrix and deducing therefrom a primitive or a near-primitive connection matrix corresponding to an optimal or at least to an economical realization of the requirements. We shall discuss these problems in the order stated, but first we need to indicate some properties of the type of Boolean matrix we use in studying switching circuits.

3. THE ALGEBRA OF SWITCHING MATRICES

3.1. *Basic Definitions and Properties*

The Boolean matrices which are useful in switching theory have all 1's on the main diagonal. Any Boolean matrix of this kind, with its remaining elements chosen from a switching algebra S, will be called a *switching matrix*. When, as in later sections of this paper, only relay contacts are used as switching elements, the resulting switching matrices are all *symmetric*. Throughout this section, we discuss the more general case, however, by way of laying the groundwork for a later extension of these methods to electronic circuits.

Consider now the set "M" of all switching matrices of order m with elements from S. We make the following definitions, where $A = [a_{ij}]$, $B = [b_{ij}]$, $\cdots$, are matrices of M:

(1) *Equality:* $A = B$ if and only if $a_{ij} = b_{ij}$ for all i and j.

(2) *Sum:* $A + B = [a_{ij} + b_{ij}]$, that is, the sum is formed by adding corresponding elements. The sum of two switching matrices is again a


$$C = \begin{bmatrix} 1 & x(y+z) & uy & w \\ x(y+z) & 1 & x+z & z \\ uy & x+z & 1 & y \\ w & z & y & 1 \end{bmatrix}$$
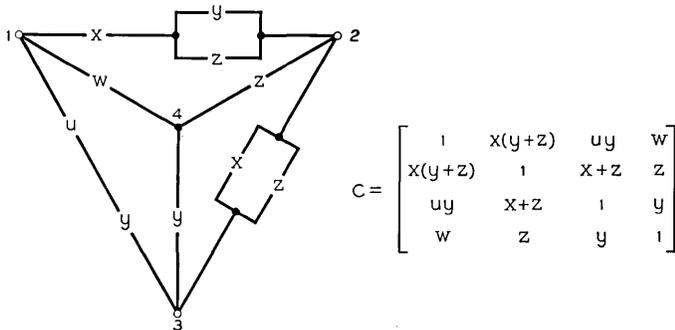
Fig. 3

switching matrix. It corresponds to connecting the elements a_{ij} and b_{ij} in parallel between nodes i and j throughout the circuit.

(3) *Logical Product:* $A * B = [a_{ij} \cdot b_{ij}]$, that is, the logical product is found by multiplying corresponding elements throughout. The logical product of two switching matrices is again a switching matrix. It corresponds to connecting the elements a_{ij} and b_{ij} in series between i and j.

(4) *Complement:* $A' = [\alpha_{ij}]$ where $\alpha_{ij} = a_{ij}'$ if $i \neq j$, but $\alpha_{ii} = 1$ for all i. This operation corresponds to replacing all the two-terminal circuits corresponding to the a_{ij} ($i \neq j$) by their complements, recognizing the fact that the connection of a terminal to itself is invariable.

(5) *Inclusion:* $A \leq B$ ("A is included in B" or "A is contained in B") if and only if $a_{ij} \leq b_{ij}$ for all i and j. Also, $B \geq A$ is equivalent to $A \leq B$. If $A \leq B$, then any combination of values of the input variables which results in a path from i to j in the circuit corresponding to A, also results in such a path in the circuit corresponding to B.

(6) *Zero Matrix:* The zero matrix Z has $a_{ij} = 0$ for $i \neq j$ but $a_{ii} = 1$ for all i. This corresponds to open circuits between all pairs of terminals.

(7) *Universal Matrix:* The universal matrix U has $a_{ij} = 1$ for all i and j. It corresponds to short circuits between all pairs of terminals.

(8) *Matrix Product:*

$$AB = \left[\left(\sum_{k=1}^{m} a_{ik} b_{kj} \right) \right].$$

The rule here is the same as for ordinary matrices. A^p means $AA \cdots A$ to p factors. The matrix product of two switching matrices is again a switching matrix, but since the product of symmetric matrices is not necessarily symmetric, this product does not always have meaning in the case of relay switching circuits.

(9) *Multiplication by a Scalar:* $\alpha A = A\alpha = [\beta_{ij}]$ where α belongs to S and $\beta_{ij} = \alpha a_{ij}$ if $i \neq j$, but $\beta_{ii} = 1$ for all i. Thus αA is again a switching matrix.

(10) *Transpose:* $A^T = [\alpha_{ij}]$ where $\alpha_{ij} = a_{ji}$.

Using these definitions, it is not difficult to prove the following fact:

Theorem 3.1.1: With respect to the meanings of $=$ and $\leq$, and the operations $+$, $*$, and $'$ as here defined, the switching matrices of order m over a switching algebra S constitute a Boolean algebra.

This result is due to Lunts.[5] As a consequence of this theorem, *we may employ all the rules of Boolean algebra in operating with switching matrices.*

When the matrix product, the transpose, and multiplication by scalars are also taken into account, many of the familiar rules of ordinary matrix algebra are seen to persist. Some additional rules result from the combina-

tion of these operations with the Boolean ones, but most of them will not be used in this paper. A list of the more fundamental properties is given in the appendix. Properties of the class of *all* Boolean matrices are discussed in Reference 6.

3.2. *A Useful Theorem*

The following result is useful in establishing several basic theorems concerning the analysis of circuits. The theorem was first stated by Lunts.[5]

Theorem 3.2.1: If A is any switching matrix of order m, then there exists a positive integer $q \leq m - 1$ such that

$$A \leq A^2 \leq \cdots \leq A^q = A^{q+1} = \cdots .$$

First we note that if $A^h = [\alpha_{ik}]$ then, since $a_{jj} = 1$,

$$A^{h+1} = \left[\sum_k \alpha_{ik} a_{kj} \right] = \left[\alpha_{ij} + \sum_{k \neq j} \alpha_{ik} a_{kj} \right].$$

Thus the ij-entry of A^{h+1} contains α_{ij} so that $A^h \leq A^{h+1}$ for all positive integers h.

To complete the proof, it will suffice to show that $A^{m-1} = A^m$. Hence consider any off-diagonal entry of A^m. (The diagonal entries are all 1.) It may be written in the form

$$\sum_{k_1 \cdots k_{m-1}} a_{ik_1} a_{k_1 k_2} \cdots a_{k_{m-2} k_{m-1}} a_{k_{m-1} j} .$$

There are $m + 1$ subscripts here, so that not all can be distinct. Consider now any term of this sum. If $j = k_s$ for some s, the term takes the form

$$a_{ik_1} \cdots a_{k_{s-1} j} a_{j k_{s+1}} \cdots a_{k_{m-1} j}$$

which is contained in the term

$$a_{ik_1} \cdots a_{k_{s-1} j}$$

of the ij-entry of A^s and hence in the ij-entry of A^{m-1}, by what has already been proved. A similar conclusion holds if $i = k_s$. If neither i nor j is equal to any k, then $k_s = k_r$ for some s and r and the term takes the form

$$a_{ik_1} \cdots a_{k_{s-1} k_r} a_{k_r k_{s+1}} \cdots a_{k_{r-1} k_r} a_{k_r k_{r+1}} \cdots a_{k_{m-1} j}$$

which is contained in the term

$$a_{ik_1} \cdots a_{k_{s-1} k_r} a_{k_r k_{r+1}} \cdots a_{k_{m-1} j} .$$

But this too is contained in the ij-entry of A^{m-1}. Thus we may conclude that $A^{m-1} \geqq A^m$. But we have already shown that $A^{m-1} \leqq A^m$. Hence $A^{m-1} = A^m$, and the theorem follows.

In the generic case, $q = m - 1$. However, because of special behavior of the elements of A, we may have $q < m - 1$, as is frequently the case when switching circuits are under consideration.

4. THE ANALYSIS OF COMBINATIONAL CIRCUITS

4.1 *The Star-Mesh Transformation and the Reduced Connection Matrix*

The basic problem of analysis is, as stated earlier, the determination of the relation between any given connection matrix and the corresponding output matrix. To accomplish this, we show first how to obtain from a given circuit an equivalent circuit using one less non-terminal node in the formation of the connection matrix. This operation may then be repeated until there are no non-terminal nodes in the accounting. The method is to formalize the Y-Δ or star-mesh transformation (Reference 3, page 94).

Consider a non-terminal node r in a combinational network with p terminal nodes and k non-terminal nodes, and with a corresponding connection matrix C, not necessarily primitive. Connections c_{ir} and c_{rj} provide a path from node i to node j if and only if $c_{ir}c_{rj} = 1$. Let us now replace the connections of the given circuit by others such that between each pair of nodes i and j (neither of which is r) there appears circuitry corresponding to the function $c_{ij} + c_{ir}c_{rj}$ and remove all connections c_{ir} between node r and other nodes of the circuit. Thus node r is effectively removed from the circuit, but the output of the circuit on the remaining nodes will be the same as before.

Matrixwise, this operation proceeds as follows. To remove a non-terminal node r, we add to each entry c_{ij} of C the product of the entry
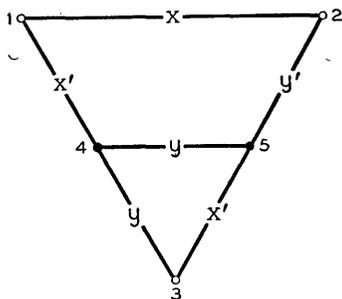
Fig. 4

c_{ir} in row i and column r of C by the entry c_{rj} in column j and row r, thereafter deleting row r and column r from C.

This operation of removing a node may be repeated until, after k steps, no non-terminal nodes remain. The resulting $p \times p$ connection matrix will be called a *reduced connection matrix* of the original circuit and will be denoted by the symbol C_0. The process is illustrated for the circuit of Fig. 4. The matrix obtained from C by the removal of node j is denoted by $C_{(j}$, etc.

$$
P = \begin{bmatrix}
1 & x & 0 & x' & 0 \\
x & 1 & 0 & 0 & y' \\
0 & 0 & 1 & y & x' \\
x' & 0 & y & 1 & y \\
0 & y' & x' & y & 1
\end{bmatrix}, \qquad
C_{(5} = \begin{bmatrix}
1 & x & 0 & x' \\
x & 1 & x'y' & 0 \\
0 & x'y' & 1 & y \\
x' & 0 & y & 1
\end{bmatrix},
$$

$$
C_0 = C_{(4(5} = \begin{bmatrix}
1 & x & x'y \\
x & 1 & x'y' \\
x'y & x'y' & 1
\end{bmatrix}.
$$

The given circuit is needlessly complicated, considering its output. It will be simplified presently.

The process of removing a node may of course be reversed in certain cases. This reversal is a simple matter when the entries of the output or connection matrix contain the proper terms. Thus, for example, we have

$$
\begin{bmatrix}
1 & a + \alpha\beta & b + \alpha\delta \\
a + \alpha\beta & 1 & c + \beta\delta \\
b + \alpha\delta & c + \beta\delta & 1
\end{bmatrix} \sim
\begin{bmatrix}
1 & a & b & \alpha \\
a & 1 & c & \beta \\
b & c & 1 & \delta \\
\alpha & \beta & \delta & 1
\end{bmatrix}.
$$

Again, starting with the matrix C_0 associated with Fig. 4, we observe that because of the common factor x', the entries $x'y$ and $x'y'$ could have arisen from the removal of a node. In fact, it is readily checked that

$$
\begin{bmatrix}
1 & x & x'y \\
x & 1 & x'y' \\
x'y & x'y' & 1
\end{bmatrix} \sim
\begin{bmatrix}
1 & x & 0 & y \\
x & 1 & 0 & y' \\
0 & 0 & 1 & x' \\
y & y' & x' & 1
\end{bmatrix}.
$$

The circuit corresponding to the second of these matrices is shown in Fig. 5. It has one less contact than the circuit with which we began, the node insertion having replaced the two contacts x' by one. As this example suggests, the "insertion of nodes" will appear to be an important tool in the synthesis of circuits.

4.2. *The Fundamental Theorem*

The relationship between connection and output matrices may now be established. (A less general form of the following theorem was first stated by Lunts.[5])

Theorem 4.2.1: If C is any connection matrix of a p-terminal circuit, if C_0 is the corresponding reduced connection matrix, and if F is the output matrix of the circuit, then there exists an integer k, $1 \leqq k < p$, such that $C_0^{p-k} = F$.

In the case of the matrix C_0 obtained in the preceding section, we have, for example, $C_0^2 = C_0$ so that C_0 is itself the output matrix of the circuit.

From Theorem 3.2.1 it follows that there exists an integer k, $1 \leqq k < p$, such that $C_0^{p-k} = C_0^{p-k+1} = \cdots$. It only remains to show that $C_0^{p-k} = F$. For this purpose it is sufficient to show that $C_0^{p-1} = F$.

Let us denote the elements of C_0 by c_{ij}. Then the ij-entry of C_0^2 is the function

$$c_{i1}c_{1j} + c_{i2}c_{2j} + \cdots + c_{ip}c_{pj} .$$

Since multiplication means "and" and addition means "or", this function is 1 for $i \neq j$ when and only when the input variables are such that there is a path from i to j, either directly (because of the term $c_{ij}c_{jj} = c_{ij}$) or via some intermediate node r. Similarly, the ij-entry of
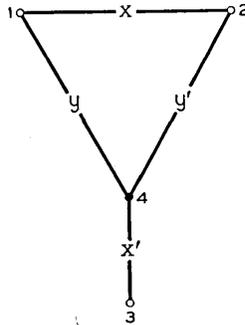
Fig. 5

$C_0{}^3$ is a function which is 1 when and only when the input variables are such that there is a path from i to j, proceeding directly, or via one intermediate node, or via two intermediate nodes. Continuing thus, since no path requires more than $p - 2$ intermediate nodes, we see that the ij-entry of $C_0{}^{p-1}$ is a function which is 1 when and only when the circuit variables are such as to interconnect i and j. That is, $C_0{}^{p-1} = F$.

The following two corollaries are immediate:

Corollary 4.2.2: The reduced circuit matrix of a two-terminal circuit is the output matrix of the circuit.

Corollary 4.2.3: The ij-entry of F may be found by considering the circuit as a two-terminal circuit connecting i and j and removing all the other nodes.

4.3. *Characterization of an Output Matrix*

Certainly any symmetric switching matrix may be interpreted as a connection matrix of a combinational relay circuit. However, a natural question to ask at this point is, "When is a given symmetric switching matrix also an output matrix?" The answer is given in

Theorem 4.3.1: The necessary and sufficient condition that a symmetric switching matrix C be an output matrix is that $C^2 = C$.

Suppose first $C^2 = C$. Multiplying both sides repeatedly by C, we conclude $C^{p-1} = C$, where p is the order of C. That is, C is its own output matrix.

Conversely, suppose C is an output matrix. Denote any reduced connection matrix of the circuit by C_0. Then, using Theorems 4.2.1 and 3.2.1, we have $C = C_0{}^{p-1} = C_0{}^{2p-2} = C^2$, and the theorem is proved.

4.4 *Redundant Elements*

In the synthesis of a circuit it is often helpful to detect and remove, or to insert, what we call *redundant elements*. These are elements whose replacement by open circuits (in the parallel case) or by short circuits (in the series case) will not alter the output of the circuit.

To illustrate these notions, we consider first the following connection matrix in which redundant terms are bracketed:

$$\begin{bmatrix} 1 & x + [y] & y & z + [xu] \\ x + [y] & 1 & y & u \\ y & y & 1 & [yuz] \\ z + [xu] & u & [yuz] & 1 \end{bmatrix}.$$

The term y in the 1,2-entry is redundant since there is a path from 1 to 3 and another from 3 to 2, hence one from 1 to 2, if $y = 1$. Thus the term y in the 1,2-position (and of course also the term y in the 2,1-position) may be dropped, that is, may be replaced by zero, and the output matrix of the circuit will not be altered. We could have reasoned in the same way that the y in the 1,3-entry or the 2,3-entry is redundant, but only *one* of these three y's may be removed. (The reader should note that the key to this situation is the appearance of two identical elements in the same row, along with a third identical element in the same column as one of them.)

Similarly, after the y in the 1,2-position has been removed, the term xu in the 1,4-entry may be seen to be redundant since there is a path from 1 to 2 if $x = 1$ and from 2 to 4 if $u = 1$. (The key here is the fact that the factors x and u of xu appear as terms of other entries in row 1 and column 4.) Finally, the entry yuz in the 3,4-position is redundant. In fact, there is a path from 3 to 2 if $y = 1$ and from 2 to 4 if $u = 1$, so that there is a path from 3 to 4 when $yu = 1$, regardless of the value of z. The successive deletion of these redundant terms, in brackets, evidently yields a primitive connection matrix which is equivalent to the original matrix in the sense that both lead to the same output matrix.

To illustrate the removal of redundant factors, we consider the connection matrix

$$
\begin{bmatrix}
1 & x & z + [x']y \\
x & 1 & x \\
z + [x']y & x & 1
\end{bmatrix}.
$$

Here the factor x' in the 1,3-entry is redundant since there is a path from 1 to 3 when y is 1, regardless of the condition of x, as a figure will readily show. This factor may therefore be dropped (replaced by 1) without altering the output matrix. This amounts to adding a redundant term x to the 1,3-entry because of the x's in the 1,2 and 2,3 positions. Then the rule $x + x'y = x + y$ accounts for the removal of x'.

The importance of these ideas is that we can reverse both processes whenever this is of avail in the synthesis of a circuit, as will appear later.

It should not be overlooked that the removal (or insertion) of redundant elements must either proceed in successive steps or be capable of being so arranged, since such an operation alters the corresponding circuit, and hence may alter the conditions for redundancy of other terms. Sometimes the insertion of redundant terms renders other terms redun-

dant so that they in turn may be removed. Later, we shall employ this fact to good advantage.

It may of course be possible to replace elements of a connection matrix by other functions than 0 and 1, but we do not employ such substitutions in this paper. These substitutions have been characterized completely by Semon.[4]

5. THE SYNTHESIS OF COMBINATIONAL CIRCUITS

5.1. *The Truth-Table Method of Synthesis*

The "truth-table" approach may be employed in synthesis in the same way that it is in the two-terminal case. For example, the output of a four-terminal circuit is specified by Table 1, where x and y are input variables.

The necessary and sufficient condition that given output requirements be consistent is clearly that whenever $f_{ij} = f_{jk} = 1$, then $f_{ik} = 1$ also. It is readily checked that this condition is satisfied in the case of this example.

We have, from the f_{12}-column, $f_{12} = x'y' + x'y = x'$. Similarly, $f_{13} = x'y + xy = y$, $f_{14} = 0$, $f_{23} = x'y$, $f_{24} = xy'$, $f_{34} = x'y'$. Thus we have

$$F = \begin{bmatrix} 1 & x' & y & 0 \\ x' & 1 & x'y & xy' \\ y & x'y & 1 & x'y' \\ 0 & xy' & x'y' & 1 \end{bmatrix}.$$

Let us consider this output matrix as a connection matrix of the desired circuit. Then, since we have a path from 2 to 1 if $x' = 1$ and from 1 to 3 if $y = 1$, the 23-entry $x'y$ is redundant and may be replaced by zero. (The resulting matrix is no longer an output matrix, of course.) Now we insert a node "5" to separate the products xy' and $x'y'$ in the

TABLE I

x	y	f_{12}	f_{13}	f_{14}	f_{23}	f_{24}	f_{34}	Non-Vanishing Product
0	0	1	0	0	0	0	1	$x'y'$
0	1	1	1	0	1	0	0	$x'y$
1	0	0	0	0	0	1	0	xy'
1	1	0	1	0	0	0	0	xy

last column:

$$\begin{bmatrix} 1 & x' & y & 0 & 0 \\ x' & 1 & 0 & 0 & x \\ y & 0 & 1 & 0 & x' \\ 0 & 0 & 0 & 1 & y' \\ 0 & x & x' & y' & 1 \end{bmatrix}.$$

The reader should check this by removing the additional node, and should *observe how this insertion replaced two y' contacts by one.* The circuit corresponding to this primitive connection matrix is shown in Fig. 6.

In many cases, the output matrix or a suitable connection matrix may be written by inspection. Any techniques useful in simplifying switching functions may of course be applied in computing the f_{ij}. However, the "simplest" forms of these functions are not necessarily the most useful, when it comes to matrix methods of synthesis. These points will be illustrated by later examples.

5.2. *Matrix Synthesis of Two-Terminal Circuits*

In the previous section, we gave an example of how one may deduce from an output matrix a primitive connection matrix of an economical realization of the circuit. In this section we illustrate the procedure further by applying it to some simple two-terminal circuits.

Geometrically, synthesis of a two-terminal circuit amounts to the selection of an appropriate set of nodes and connecting links joining the
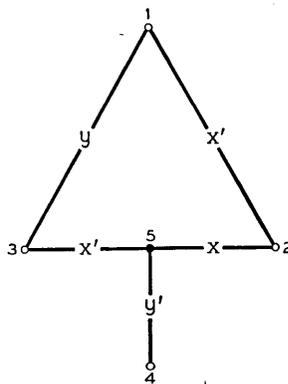
Fig. 6

two terminals. Algebraically, this is accomplished by beginning with the output matrix

$$\begin{bmatrix} 1 & f \\ f & 1 \end{bmatrix}$$

of the two-terminal circuit and reducing the complexity of its entries by successive node insertions until a primitive connection matrix is finally obtained. Operations with redundant elements are ordinarily an essential part of the process, as the following examples show.

Example A. $f = ABC + AD + BD + CD$.

First we factor f in some convenient way, say into the form $f = A(BC + D) + BD + CD$, and by the insertion of a node remove the first term, that is, render it redundant so that it may be replaced by zero:

$$\begin{bmatrix} 1 & A(BC + D) + BD + CD \\ A(BC + D) + BD + CD & 1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & BD + CD & A \\ BD + CD & 1 & BC + D \\ A & BC + D & 1 \end{bmatrix}.$$

Suppose now that we decide to remove the term CD from the 1,2-entry. We note first that there is a path from 1 to 2 if $BD = 1$ and from 2 to 3 if $D = 1$, i.e., from 1 to 3 if $BD = 1$. Hence we may add the redundant term BD to the 1,3-entry:

$$\begin{bmatrix} 1 & BD + CD & A + [BD] \\ BD + CD & 1 & BC + D \\ A + [BD] & BC + D & 1 \end{bmatrix}.$$

(Brackets around a term denote that it is redundant.) Now we can insert a fourth node which removes the terms CD, BD, BC from the 1,2-, 1,3-, and 2,3-entries respectively. This yields the matrix

$$\begin{bmatrix} 1 & [BD] & A & D \\ [BD] & 1 & D & C \\ A & D & 1 & B \\ D & C & B & 1 \end{bmatrix}.$$

Here the entry BD is redundant, for we have a path from 1 to 4 if $D = 1$, from 4 to 3 if $B = 1$, from 3 to 2 if $D = 1$, hence from 1 to 2 if $BD = 1$. Dropping the two BD's we obtain the primitive connection matrix,

$$\begin{bmatrix} 1 & 0 & A & D \\ 0 & 1 & D & C \\ A & D & 1 & B \\ D & C & B & 1 \end{bmatrix}$$

which is a wiring diagram for the bridge circuit shown in Fig. 7.

This work can all be performed without recopying. Thus, the matrix

$$\begin{bmatrix} 1 & [A(BC+D)]+[BD]+[CD] & A+[BD] & D \\ [A(BC+D)]+[BD]+[CD] & 1 & [BC]+D & C \\ A+[BD] & [BC]+D & 1 & B \\ D & C & B & 1 \end{bmatrix}$$

gives the desired result. Brackets are drawn around all terms, whether originally present or inserted, which are ultimately removed because of redundancy.

Example B. $f = A'B + AB' + AC$.

One possible procedure is indicated by the following matrix:

$$\begin{bmatrix} 1 & [A'B]+[AB']+[AC] & A & A & A' \\ [A'B]+[AB']+[AC] & 1 & C & B' & B \\ A & C & 1 & 0 & 0 \\ A & B' & 0 & 1 & 0 \\ A' & B & 0 & 0 & 1 \end{bmatrix}.$$

Here the three terms of f were removed one at a time, proceeding from right to left, by inserting three additional nodes. The corresponding circuit, shown in Fig. 8, contains an unnecessary A-contact, even though we have arrived at a primitive connection matrix.

An alternative procedure is indicated in the matrix:

$$\begin{bmatrix} 1 & [A'B]+[A(B'+C)] & A' & A \\ [A'B]+[A(B'+C)] & 1 & B & B'+C \\ A' & B & 1 & 0 \\ A & B'+C & 0 & 1 \end{bmatrix}$$

in which the $A'B$ term was removed first. The corresponding, more economical circuit is given in Fig. 9.

These examples illustrate the important facts that (1) *the matrix representation is not prejudiced in favor of series-parallel circuitry* and (2) *the circuit finally obtained depends on the steps used in obtaining a primitive connection matrix.*

5.3. *Further Examples With More Than Two Terminals*

The examples of the preceding section were introduced primarily for illustrative purposes. We now introduce two examples designed to indicate the power of the method.

First we construct a circuit simultaneously realizing all sixteen switching functions of two variables. (This circuit was first obtained by a
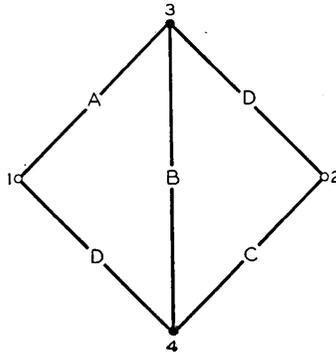
Fig. 7
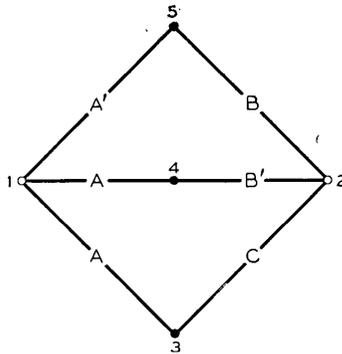
Fig. 8

student in the switching course at M.I.T. and was later proved by Shannon, in an unpublished memorandum, to be minimal.)

A connection matrix of order 17 (sixteen terminals and ground) may be written at once. It has the 16 functions in the G row and column, and all other off-diagonal entries are zero. For convenience, we omit

1 2 3 4 5 6 7 8 9 10 11 12 13	14	15	16	G	
1				0	1
1	y' y			x	2
1				y	3
1	y' y			x'	4
1				y'	5
1 x'	$x + y$		$[x'] + y'$	6
1	$x + y'$		$[x'] + y$	7
1 x			$[x] + y'$	8
1			$[x] + y$	9
1			$[x'y']$	10
1			$[x'y]$	11
1			$[xy']$	12
1			$[xy]$	13
1		$[(x' + y)(x + y')]$	14
1	$[(x' + y')(x + y)]$	15
1	1	16
	1	G

all entries below the diagonal. Entries which are zero throughout are omitted entirely. (Off-diagonal entries which are not in the G-column are inserted redundant terms whose presence will be explained.)
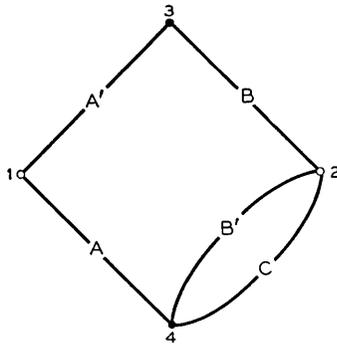
Fig. 9

First we note that since terminals 4, 10, and 11 are grounded when and only when $x' = 1$, $x'y' = 1$, and $x'y = 1$ respectively, redundant entries y' and y may be inserted in the 4,10- and 4,11-positions respectively. Then, however, the 10,G- and 11,G-entries become redundant and may be replaced by zeros. (The terms y and y' which we inserted then lose their redundancy, of course.) In the same way, we may put redundant terms y' and y in the 2,12- and 2,13-entries respectively, thereafter replacing the now redundant 12,G- and 13,G-entries by zeros.

With the entries $x'y' + xy$ and $x'y + xy'$ in the 14,G- and 15,G-positions factored as indicated in the matrix, we see next that the same type of operation permits insertion of redundant entires $x + y'$ and $x + y$ in the 7,14- and 6,15-positions respectively, after which the 14,G- and 15,G-entries become redundant and may be removed.

Finally, we note that the common term x' in the 6,G- and 7,G-entries permits the insertion of a redundant x' in the 6,7-position. This in turn renders the x''s in the 6,G- and 7,G-entries redundant, so that they may be replaced by zeros. Finally, in the same way, we insert a redundant x in the 8,9-position, after which the x's in the 8,G- and 9,G-entries may be replaced by zeros.

The resulting primitive connection matrix corresponds to the circuit shown in Fig. 10.
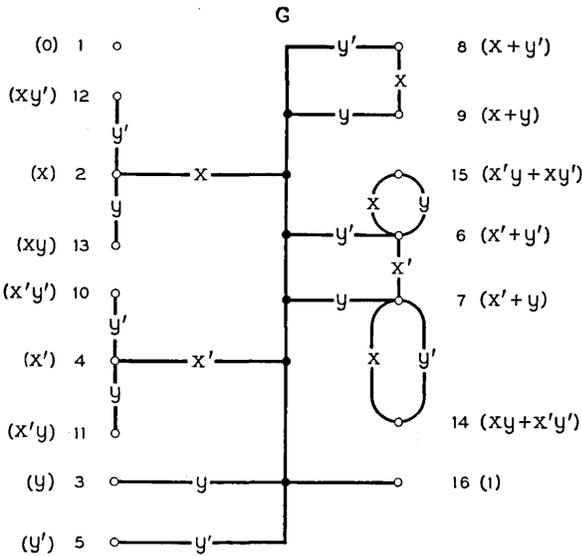
Fig. 10

TABLE II

Relays Operated	Leads Grounded
v, w	1
v, x	2
w, x	1, 2
v, y	3
w, y	1, 3
x, y	2, 3
v, z	1, 2, 3
w, z	4
x, z	1, 4
y, z	2, 4
None	None

As a second example, consider a circuit (Reference 3, page 124) with five relays, v, w, x, y, z and leads 1, 2, 3, 4, whose operate conditions are given in Table II.

No other combinations of relays operated occur, so that we "don't care" what happens in the circuit for such combinations. Nor do we care if ungrounded terminals are interconnected. Taking account of these assumptions, it may now be seen that switching functions expressing the conditions under which the various leads are grounded are as shown in Table III.

These functions allow us to write a connection matrix for the desired circuit:

$$
\begin{array}{ccccc}
(1) & (2) & (3) & (4) & (G)
\end{array}
$$

$$
\begin{bmatrix}
1 & 0 & 0 & 0 & wz' + w'zy' \\
0 & 1 & 0 & 0 & xz' + x'zw' \\
0 & 0 & 1 & 0 & vz + yz' \\
0 & 0 & 0 & 1 & zv' \\
wz' + w'zy' & xz' + x'zw' & vz + yz' & zv' & 1
\end{bmatrix}.
$$

The insertion of a node "5" to remove the terms in column "G" which contain the factor z immediately suggests itself since 4 z-contacts might thus be replaced by a single z-contact. This would require, however, preliminary insertion of suitable redundant terms in place of certain zero-entries of this matrix. The additional column and row for node "5" and the requisite terms whose redundancy must be investigated in the light of the don't-care conditions are shown in the next matrix.
(Proposed redundant terms are listed above the diagonal only.)

TABLE III

Lead	Function
(1)	$wz' + w'zy'$
(2)	$xz' + x'zw'$
(3)	$vz + yz'$
(4)	zv'

In the 1,2-position, the product $x'w'y'$ would have to be inserted, as shown. But when $x'w'y' = 1$, vz may be 1 also. May 1 and 2 be con-

$$\begin{bmatrix} \begin{bmatrix} 1 & (x'w'y') & (vw'y') & (v'w'y') & wz' + w'zy' \\ 0 & 1 & (vw'x') & (v'w'x') & xz' + x'zw' \\ 0 & 0 & 1 & 0 & vz + yz' \\ 0 & 0 & 0 & 1 & zv' \\ wz' + w'zy' & xz' + x'zw' & vz + yz' & zv' & 1 \end{bmatrix} & \begin{matrix} w'y' \\ w'x' \\ v \\ v' \\ z \end{matrix} \\ \begin{matrix} w'y' & w'x' & v & v' & z \end{matrix} & 1 \end{bmatrix}$$

nected when $vz = 1$? A check of the table of operate-conditions shows that 1 and 2 are both to be grounded when $vz = 1$, so that the insertion of this term is harmless.

Next, in the 1,3-position, the entry $vw'y'$ would have to be inserted. When this factor is 1, we may also have $xz = 1$, but v and x and z are never all simultaneously operated, so this causes no trouble. However, we may alternatively have $xz' = 1$ or $x'z = 1$. May 1 and 3 be connected when v and x or v and z are both operated? The table shows that, in either case, 1 and 3 may be connected since neither is grounded in the first case, but both are grounded in the second. Thus the term $vw'y'$ may be safely inserted.

The term $vw'x'$ in the 2,3-position brings trouble, however, for when $vy = 1$, only 3 is to be grounded, whereas 2 and 3 could be connected in this case. If we abandon the attempt to remove the term vz from the $3,G$-entry at this step, the difficulty is eliminated, for the required redundant terms then violate none of the operate conditions of the circuit.

The terms containing z' may also be removed from column G by the insertion of a node "6". This replaces three z-contacts by just one. The insertion of both nodes 5 and 6 is indicated in the following matrix. It is left to the reader to complete the checking of the redundant terms.

The redundant terms we inserted were of course absorbed again by the node-insertion process. We now have a connection matrix which

$$
\begin{array}{ccccccc}
(1) & (2) & (3) & (4) & (G) & (5) & (6) \\
1 & 0 & 0 & 0 & [wz'] + [w'zy'] & w'y' & w \\
0 & 1 & 0 & 0 & [xz'] + [x'zw'] & w'x' & x \\
0 & 0 & 1 & 0 & vz + [yz'] & 0 & y \\
0 & 0 & 0 & 1 & [zv'] & v' & 0 \\
[wz'] + [w'zy'] & [xz'] + [x'zw'] & vz + [yz'] & [zv'] & 1 & z & z' \\
w'y' & w'x' & 0 & v' & z & 1 & 0 \\
w & x & y & 0 & z' & 0 & 1
\end{array}
$$

is not primitive but which cannot be further simplified by node insertion because of the absence of appropriate common factors. The corresponding circuit is shown in Fig. 11.

The reader may check that the requirements are satisfied and that no leads are improperly grounded. Although there are 12 contacts in this realization, only 21 springs are required because of the three possible transfers. Leads 1, 2, and 4 are connected when none of the relays are operated, but otherwise no ungrounded leads are connected.

5.4. *Other Transformations of a Connection Matrix*

We have seen how the removal and insertion of nodes by the Y-Δ transformation may be used in the analysis and synthesis of networks.
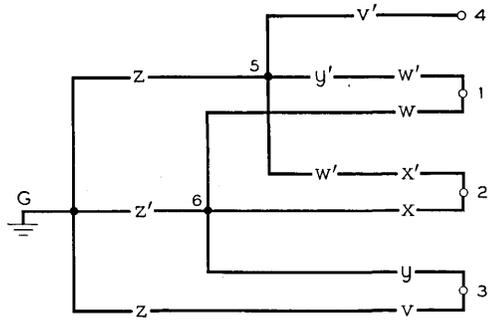

Fig. 11

There are, of course, other methods of transforming a connection matrix without altering the output. One of these is the Δ-Y transformation, which is simply the dual of the one just mentioned, and which we now explain.

Consider three nodes i, j, k of a network, as indicated in Fig. 12(a). This part of the network may be replaced by the network of Figure 12(b). This replacement is what is known as the "Δ-Y transformation" (Reference 3, page 93).

Matrixwise, this transformation is simple to execute. We first mark the ij and ji, jk and kj, ki and ik entries of the connection matrix, say by bracketing them. Then, in a new column, we enter in rows i, j, k the sums of the marked entries in those rows. The rest of the column is filled out with zeros except for the diagonal entry, which — as always — is 1. The bracketed elements are then replaced by zeros and the matrix is completed in symmetric fashion. The reduced connection matrix appearing in Section 4.1 is used to provide an example:

$$
\begin{bmatrix}
1 & [x] & [x'y] \\
[x] & 1 & [x'y'] \\
[x'y] & [x'y'] & 1
\end{bmatrix}
\sim
\begin{bmatrix}
1 & 0 & 0 & x+y \\
0 & 1 & 0 & x+y' \\
0 & 0 & 1 & x' \\
x+y & x+y' & x' & 1
\end{bmatrix}.
$$

The result is a primitive connection matrix, but not as simple a one as we had before. However, a redundant x may be inserted in the 1,2-



(a)   (b)
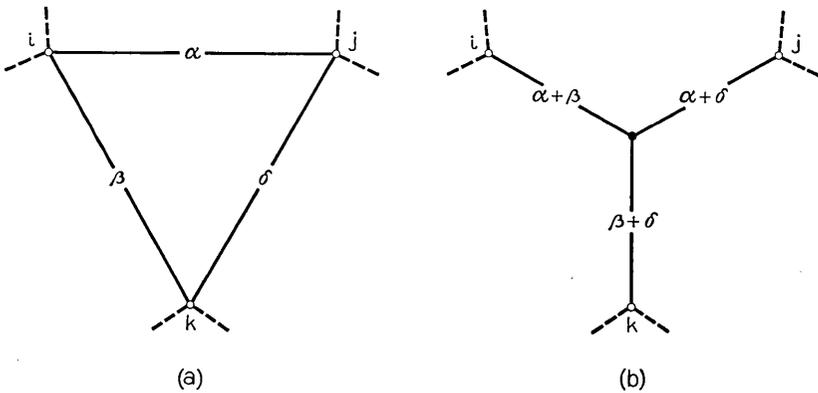
Fig. 12

position after which the x's in the 1,4- and 2,4-positions become redundant. When these are removed, the matrix is again that of the circuit in Fig. 5.

This transformation is clearly indicated for consideration when the addition of two entries in the same row results in a considerable simplification. It is probable that still other transformations — useful, like this one, under special conditions — could be developed. However, because of its simple geometric and algebraic significance, the node-insertion operation seems likely to remain most useful of all.

6. CONCLUSION

In this paper we have outlined the basic properties of switching matrices and have applied them to both the analysis and the synthesis of combinational relay circuits. It is clear that we have not reduced circuit design to a "cook-book" procedure, but our experience with a variety of design problems (not all of which are reported here) leads us to believe that the method shows considerable promise of becoming a "practical" tool, and that further study is justified along the following lines:

(a) The work done in References 4 and 7 on circuits with unilateral elements should be extended in the hope of devising a tool comparable to node-insertion for synthesis.

(b) The class of all transformations of a connection matrix which leave the associated output matrix invariant should be characterized and their application in synthesis should be studied. (From an algebraic point of view, the interesting thing here is that allowable transformations of a connection matrix need not leave its order invariant.)

(c) The manner in which "don't-care" conditions enter into synthesis should be more extensively studied. In the presence of such conditions, the class of connection matrices giving rise to an acceptable output matrix is of course considerably more extensive than it would be otherwise.

(d) The possibility of characterizing a primitive connection matrix of a minimal network should be investigated. This may be related to synthesis by a minimum number of nodes and complete absence of redundant elements.

Other problems have, of course, suggested themselves. We have listed what seem to be the more important ones; the reader will undoubtedly formulate others for himself.

7. APPENDIX

The Basic Properties of Switching Matrices

If A, B, C are any switching matrices of order $m \cdot$ over a switching algebra with n variables, then with respect to the definitions given in Section 3, we have the following Boolean rules of operation:

$$A + A = A \qquad\qquad\qquad A * A' = Z$$
$$A * A = A \qquad\qquad\qquad U + A = U$$
$$A + B = B + A \qquad\qquad\qquad U * A = A$$
$$A * B = B * A \qquad\qquad\qquad A + A' = U$$
$$A + (B + C) = (A + B) + C \qquad\qquad (A * B)' = A' + B'$$
$$A * (B * C) = (A * B) * C \qquad\qquad (A + B)' = A' * B'$$
$$A + B * C = (A + B) * (A + C) \qquad\qquad (A')' = A$$
$$A * (B + C) = A * B + A * C \qquad\qquad A + A * B = A$$
$$Z + A = A \qquad\qquad\qquad A + A' * B = A + B$$
$$Z * A = Z \qquad\qquad\qquad\qquad A \leqq A$$

$$A \leqq B \text{ and } B \leqq A \text{ if and only if } A = B$$
$$A \leqq B \text{ and } B \leqq C \text{ imply } A \leqq C$$
$$A \leqq B \text{ if and only if } A * B = A$$
$$A \leqq B \text{ if and only if } A + B = B$$
$$Z \leqq A \leqq U \text{ for all } A.$$

Every Boolean matrix C has a canonical expansion

$$C = \sum_{k=0}^{2^n-1} C_k p_k$$

and a dual canonical expansion

$$C = \prod_{\substack{* \\ k=0}}^{2^n-1} (\Gamma_k + s_k U)$$

where p_k are the fundamental products formed from $x_1, x_2, \cdots, x_n$ and the s_k are the fundamental sums. The ij-entries of C_k and Γ_k are the values associated with p_k and s_k respectively in the ordinary canonical expansions of the entry c_{ij} of C.

When the not-characteristically-Boolean operations of forming the transpose and the matrix product are introduced, we find that the following properties hold, among others. Many are familiar, others are not.

$$AB \neq BA \text{ ordinarily} \qquad\qquad U^p = U$$
$$A(B + C) = AB + AC \qquad\qquad Z^p = Z$$
$$(A + B)C = AC + BC \qquad\qquad (A^p)^q = A^{pq}$$
$$AZ = ZA = A \qquad\qquad A^p A^q = A^{p+q}$$
$$(A^T)^T = A \qquad\qquad AU = UA = U$$
$$(A^T)' = (A')^T \qquad\qquad (AB)C = A(BC)$$
$$(A + B)^T = A^T + B^T \qquad\qquad A(B * C) \leq AB * AC$$
$$(A * B)^T = A^T * B^T \qquad\qquad (A * B)C \leq AC * BC$$
$$(AB)^T = B^T A^T$$

$A \leq B$ implies $AC \leq BC$ and $CA \leq CB$, but not conversely.

8. ACKNOWLEDGMENT

9. REFERENCES

1. C. E. Shannon, A Symbolic Analysis of Relay and Switching Circuits, Trans. A.I.E.E., **57,** 1938, pp. 713–723.
2. C. E. Shannon, The Synthesis of Two-Terminal Switching Circuits, B.S.T.J., **28,** 1949, pp. 59–98.
3. W. Keister, A. E. Ritchie, S. H. Washburn, The Design of Switching Circuits, New York, Van Nostrand, 1951.
4. W. Semon, Matrix Theory of Switching Networks (a thesis), Harvard University, The Computation Laboratory, 1954.
5. A. G. Lunts, The Application of Boolean Matrix Algebra to The Analysis and Synthesis of Relay Contact Networks (in Russian), Doklady Akademii Nauk SSSR, **70,** 1950, pp. 421–23.
6. R. D. Luce, Note on Boolean Matrix Theory, Proc. Amer. Math. Soc., **3,** 1952, pp. 382–388.
7. B. I. Aranovich, The Use of Matrix Methods in Problems of the Structural Analysis of Relay-Contact Networks (in Russian), Automatika i Telemekhanika, **10,** 1949, pp. 437–51.

Arcing of Electrical Contacts in Telephone Switching Systems

Part IV — Mechanism of the Initiation of the Short Arc

By M. M. ATALLA

A mechanism is presented for the initiation of the short arc. Breakdown in short gaps, occurring at measured fields of 30×10^6 (+ 0 per cent, − 15 per cent) volts/cm for clean palladium and 28×10^6 (+ 0 per cent, − 15 per cent) volts/cm for clean gold, are generally preceded by field emission currents from the cathode sufficient to cause heating and evaporation of the anode. This is followed by ionization of the metal vapor, enhancement of the cathode emission, etc.

(1) It is shown that to evaporate the anode by electron bombardment one must satisfy the requirement $j_-Va/k = (\Delta T)_b$. By combining this requirement with the field emission equation one may calculate the relation between arc initiation voltage and contact separation from the physical properties of the contact metal. These calculations are in agreement with the measurements.

(2) A breakdown is obtained only when a minimum power $IV = \pi a k(\Delta T)_b$ is dissipated at the anode by electron bombardment. Measurements with clean palladium contacts have given $IV = 1$ watt in agreement with the above relation. This concept of minimum power for arc initiation should replace that of minimum current for arc initiation. It is possible at 100 volts to initiate a breakdown between palladium contacts at 0.01 ampere which is 100 times smaller than the minimum current required for maintaining the arc.

(3) A breakdown will lead to an established arc only if the circuit can furnish enough current to maintain the arc plasma by sufficient metal evaporation; otherwise, transient unstable breakdowns are obtained.

(4) Breakdown time lags have been observed. They correspond approximately to the heating time of the anode by electron bombardment until

evaporation occurs. It was shown that this time lag is theoretically inversely proportional to the square of the over-voltage.

(5) By limiting the circuit currents, it was possible to exceed the usual arc initiation field without breakdown. For palladium the field could be increased to 36 × 10^6 volts/cm when a metal bridge formed electrostatically across the gap. The significance of electrostatic bridging in enhancing breakdowns, particularly in the presence of emission currents and their heating effects, is discussed.

INTRODUCTION

It has been shown[1] that an arc may be initiated between contacts at close separations at voltages well below the spark breakdown potential of the surrounding atmosphere. These arcs were found[2] to be initiated at approximately a constant field of the order of millions of volts/cm. The magnitude of the field is dependent primarily on the contact surface conditions. Essentially the same results were obtained in vacuum at small separations.[3] In this paper, arc initiation measurements were extended to a lower range of separation between 200 and 2000Å for palladium and gold contacts in air. Approximately constant field lines were still obtained for arc initiation with fields as high as 30 × 10^6 volts/cm. These are the highest observed fields for arc initiation. The corresponding emission currents are appreciable and are shown to be responsible for setting off the chain of events leading to establishment of the arc.

NOTATION

a	Effective radius of electron field emission beam from cathode from a point on cathode surface
d	Separation between the contacts
j_-	Electron current density
k	Thermal conductivity
ℓ	Height of surface irregularity on cathode surface
m	Mass of metal atom
r	Radius of curvature of a point on the cathode surface
s	Electrostatic stress
t	Time
u	Dimensionless parameter in heat conduction equations $$u = a/2(\alpha t)^{1/2}$$
$f(y)$	Nordheim elliptic function
z	Distance from the surface of the anode in the metal
F	True field strength
I	Electron current

I_i Minimum electron current to initiate a breakdown
R Circuit resistance
T Temperature
T_b Boiling temperature
T_o Ambient temperature
ΔT Temperature rise
$(\Delta T)_b$ Temperature rise to boiling
$(\Delta T)_c$ Critical temperature rise leading to a breakdown
V Voltage
V_{ai} Arc initiation voltage
V_{eb} Electrostatic bridging voltage
V_o Initial voltage
V_c Voltage across the contact
α Thermal diffusivity
φ Work function
ξ Field intensification
ρ Electric resistivity

OUTLINE

Consider a pair of contacts at a certain fixed separation. The separation here is defined as the minimum distance one contact has to move before physical contact first takes place. Due to surface irregularities one must consider this contact separation to be between two points one on each contact. Fig. 1 (a) shows a point on the cathode with a radius of curvature r and another on the anode with a much larger radius of curvature. By applying a voltage across the contacts a field distribution is obtained on the cathode point, Fig. 1 (b). This distribution is primarily dependent on the geometry involved and, in general, the maximum field is obtained at the tip of the point. If the fields obtained are high enough, perceptible field emission currents will be obtained at the cathode surface. Due to the extreme sensitivity of field emission to relatively small changes in the field, the corresponding distribution of current density on the cathode point approaches a rectangular distribution, Fig. 1 (c). Electrons emitted from the cathode are accelerated by the field in the gap and will bombard an area on the anode causing local heating, Fig. 1 (d). The maximum temperature obtained is a function of the thermal conductivity of the anode material and the size and energy of the field emission electron beam. Should this maximum temperature reach the boiling temperature* of the anode material, evaporation will

---

* It will be shown that heating to the softening temperature of the metal may be sufficient. Intensification of the field is then produced by electrostatic pulling of the anode metal.
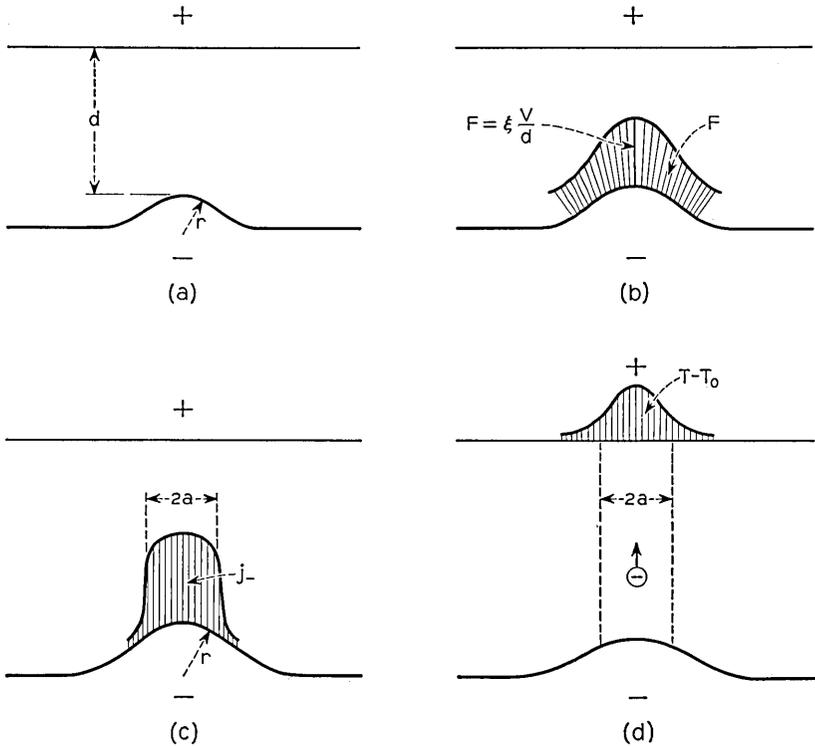
Fig. 1 — Schematic representations of contact geometry, electrostatic field, field emission current and anode heating by electron bombardment.

take place. This is the first step leading to the build-up of the arc plasma in the gap. If the potential drop is above the minimum ionization potential of the metal vapor, positive ions will be produced by electron-metal atom collision. The first ion front produced will travel towards the cathode with a cross-section greater than that of the initiating electron beam. This is due to the initial velocity of the metal atoms corresponding to the boiling temperature of the anode metal. The ion front streaming towards the cathode will increase the field at the cathode. This will multiply the original electron emission by both increasing the emission from areas already emitting and by introducing new emitting areas. This process will repeat and if the associated circuit conditions satisfy certain requirements the plasma is fully developed and the arc is established. Otherwise, a transient unstable breakdown takes place.

ARC INITIATION VOLTAGE VERSUS SEPARATION

In Reference 2 measurements of arc initiation voltage versus contact separation were given. It was concluded that breakdown fields of only a few million volts/cm were due to surface contamination. In the same reference it was indicated that when the contacts were cleaned by heavy arcing, fields as high as 20×10^{6} volts/cm failed to initiate the arc. With the range of contact separation then available it was not possible to further increase the field without obtaining spark breakdowns along longer paths. This result was obtained for the metals Ag, Ni, Pd, Pt and W.

To avoid the above difficulty a similar cantilever bar set-up was built with a contact separation resolution as low as 200Å. The accuracy of the separation setting is determined by the reproducibility of the zero point, i.e., the point at which physical contact first takes place. Best results were obtained by eliminating the use of direct current as a detector. Instead, the contacts were shunted by a pair of wires a few feet long which were directly connected to an oscilloscope through a high gain amplifier. The electromagnetic pickup of the wires is detected on the oscilloscope when the contacts are open and vanishes when the contacts are touching. When the contacts are clean it is possible to reproduce the same zero point to within less than half a division on the vernier of the micrometer. The accuracy of this setup was estimated at better than ±60Å.

The contacts were cleaned by heavy arcing as described in Reference 2. The gap was set at a certain value and the voltage gradually increased. Breakdowns were detected on an oscilloscope. It was observed that in some cases rapid transient breakdowns might occur without being detected. These breakdowns were found to change the separation and the subsequent breakdown voltage obtained did not correspond to the original separation setting. This difficulty was overcome by using a different procedure demonstrated in Table I. In this case the desired separation

TABLE I — PROCEDURE FOR DETERMINING ARC INITIATION VOLTAGE
AT A FIXED SEPARATION
Separation $d = 500$Å

Voltage level approached	110	120	135	150
Separation measured after reaching above voltage level d in 100 Å	5 5 5 5 5	5 5 5 5 5	12* 20* 5 5 5	
Arc initiation voltage ...				150 140 140 130

* Change in separation due to an undetected breakdown.
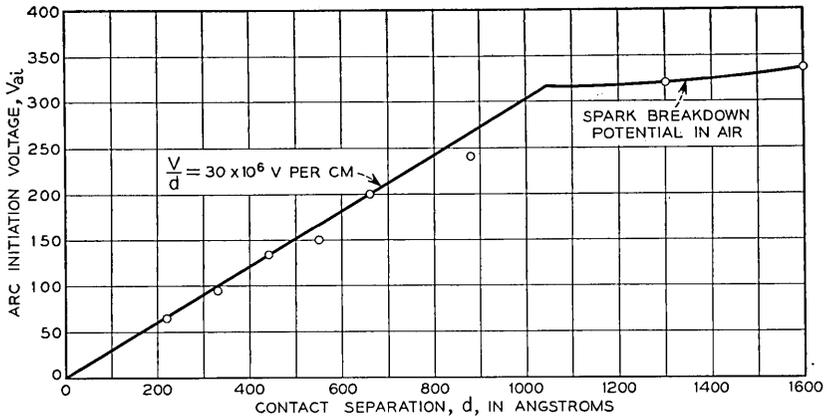
Fig. 2 — Arc initiation data for clean palladium contacts in air at low separations.

was 500Å. The voltage was gradually increased to a specific voltage level of 110 volts. If reached without an observed breakdown the separation was remeasured. The table shows that after applying this voltage no change of separation occurred. This indicated that the arc initiation was above 110 volts. Repeating at 120 volts the results indicated that the arc initiation voltage still was not exceeded.

At 135 volts the first two measurements showed an increase in separation from an initial setting of 500Å to 1,200Å and 2,000Å. The subsequent measurements gave no change. From this one concludes that the arc initiation voltage is in the neighborhood of 135 volts. Smaller voltage steps were then used. For illustration, the 150-volt level is shown where breakdowns were consistently obtained at or below 150 volts. By this process the boundaries of the arc initiation voltage are obtained and the average and spread are determined.

The results obtained for Pd contacts in air are shown in Fig. 2. For separations below 1,000Å the arcs were initiated at a value of V/d, the apparent or gross field, of 30×10^6 volts/cm. At higher separations breakdowns occurred along longer paths at the minimum sparking potential of air. Similar measurements were made for clean gold contacts. The value of V/d obtained was about 28×10^6 volts/cm. These results are in accordance with the data of Reference 2 where it was shown that for all the metals tried the value of V/d at breakdown should be greater than 20×10^6 volts/cm. These fields are appreciably higher than has been reported for these short gaps.*

* P. Kisliuk[3] has recently reported similar high fields in vacuum.

In an attempt to explore the subsequent process in the arc initiation mechanism, it was necessary to investigate the effects of the associated electrostatic stresses particularly at the high fields reported above.

ELECTROSTATIC PULLING OF METAL BRIDGES AT SMALL CONTACT SEP-ARATIONS

In this section an attempt is made to (1) establish that a breakdown cannot be produced without sufficient emission currents, and (2) determine the possible effects of the electrostatic stresses associated with the high fields leading to the breakdown.

For the measurements presented above, the circuits used were essentially low resistance circuits with little limitation on the currents they supplied. At the high fields obtained the field emission currents were appreciable. As will be shown later, these currents can cause heating of the anode or cathode leading to the subsequent steps of arc initiation. An experiment was then performed to investigate the effect on breakdown of limiting the circuit current by a high circuit resistance. A resistance of 1.1×10^7 ohms was placed in the circuit. Emission of discharge currents during the application of voltage were observed by a microammeter in the contact circuit.

The same procedure used for the determination of the arc initiation voltage was applied. At a fixed separation a critical voltage was reached at which a rapid closure of the contact gap was observed. This voltage was found to be quite sensitive to the surface conditions of the contacts. Loose particles on the surface always produced closures at appreciably lower voltages.* The results for carefully and frequently cleaned contacts were fairly consistent and reproducible. The results for Pd contacts are given in Fig. 3. For voltages below 300, emission currents were observed but were less than 2×10^{-8} ampere and the correction for the voltage drop across the circuit resistance was negligible. For larger separations higher fields could not be applied across the contact due to the establishment of glow discharge along a longer path. The voltage drop across the resistor was such that the voltage across the contacts was consistently maintained at about 320 volts.

The above data indicates that with the high resistance in the circuit the usual breakdown voltage, at a given separation, could be exceeded without breakdown up to a higher value at which sudden closure was obtained. This occurred at approximately a constant value of V/d of

---

* In extreme cases closures were observed at voltages below 10 volts for badly contaminated contact surfaces.
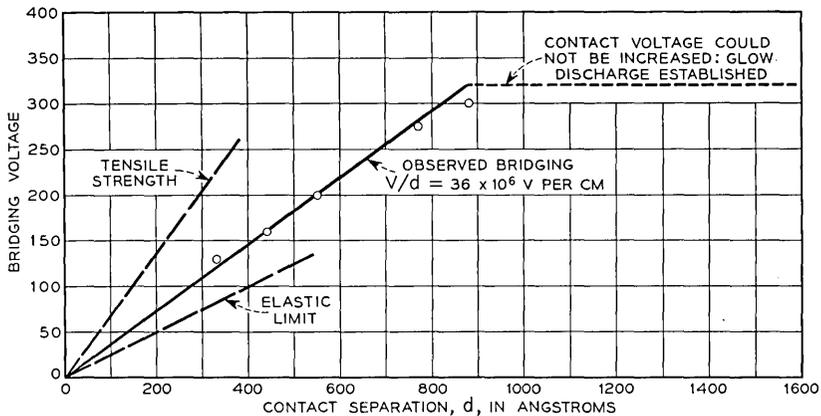
Fig. 3 — Electrostatic bridging of clean palladium contacts at small separations. Bridging voltages are corrected for potential drop across circuit resistor.

36×10^6 volts/cm.* This field could not have possibly been maintained at the contacts since the corresponding emission currents, calculated from equation (2), could not be furnished by the circuit due to the high resistance used. One, therefore, suspects that the local contact circuit, involving the contacts and a few millimeters of lead wire, were continuously discharging by field emission and recharging through the high circuit resistance as evidenced by fluctuations in the circuit current. By measurement, the current supplied by the circuit during this period was less than 2×10^{-8} ampere. The corresponding average power of these discharges was less than $300 \times 2 \times 10^{-8} = 6 \times 10^{-6}$ watts since the voltages applied were below 300 volts. This power dissipation, as shown in a later section, is not sufficient to cause any significant change in the anode temperature that can account for the closures observed. This led to the following consideration of the effect of electrostatic pulling at the high fields reported.

The surface electrostatic stress due to a field V/d may be calculated from the following equation for parallel plates with a dielectric constant 1.0:

$$s = 4.42 \times 10^{-7}(V/d)^2 \tag{1}$$

where s is in dynes/cm^2 and V/d is in volts/cm. At the observed field V/d of 36×10^6 volts/cm, the surface stress is 5.7×10^8 dynes/cm^2. Unfortunately, one cannot specify with certainty the yield strength of

* Pearson[4] has previously obtained a constant field for gold contacts. In his work, however, the circuit currents were not limited and his measurements probably correspond to our data on arc initiation.

the metal that corresponds to these stressing conditions and dimensions. For *bulk* palladium[5] the elastic limit is about 3.2×10^8 dynes/cm² and the tensile strength is about 21×10^8 dynes/cm². The corresponding fields, calculated from equation (1), are 26×10^6 and 69×10^6 volts per cm respectively, Fig. 3. This agreement would tend to substantiate the postulate that the observed closures were due to electrostatic pulling. This, however, is only true if the elastic and plastic behavior of the contact metal is similar to that of bulk metals. For these, the range of elastic strain is relatively small and this behavior is usually explained in terms of motions of imperfections, especially dislocations. On the other hand, for specimens that are small enough to be free of dislocations or have only a few, the elastic range may be much larger.[6,7]

The significance of electrostatic pulling in the initiation of the short arc may be summarized as follows. For the very soft or liquid metals, electrostatic pulling will precede the occurrence of high emission currents. This will cause an increase in the emission currents by intensifying the local fields. The arcing for these metals is essentially triggered by electrostatic pulling. For metals like palladium, on the other hand, field emission currents start the breakdown. In these cases, however, one should not preclude the significance of electrostatic pulling. It will be shown later, that due to the heating effects of the emission currents one may obtain at least partial bridging of the contact separation at lower fields than needed for the cold metal. This can cause an appreciable intensification of the local fields, the emission currents, etc. These effects are discussed in more detail in the following section.

DERIVATION OF RELATION BETWEEN ARC INITIATION VOLTAGE AND CONTACT SEPARATION

Consider the contact arrangement of Fig. 1 (d). When a potential V is suddenly applied, a field distribution on the cathode surface is obtained, $F = \xi V/d$, with a maximum value at the tip of the cathode point, where ξ is a field intensification factor which depends on the geometry and separation involved.

For illustration purposes only, Fig. 4 shows the calculated distribution of ξ over a two-dimensional surface. The corresponding distribution of current densities is also shown. This was obtained from the field emission equation:[8]

$$j_- = 1.54 \times 10^{-6} \frac{F^2}{\varphi} \exp\left[-6.83 \times 10^7 \, \varphi^{3/2} f(y)/F \right] \qquad (2)$$

where $f(y)$ is the Nordheim elliptic function[9] of the variable $y = 3.79$

$\times 10^{-4} F^{1/2}/\varphi$. A convenient tabulation of equation (2) is given in Reference 10. Due to the sensitivity of the emission to changes in the field, the current density distribution is generally much sharper than the field distribution. In the following analysis, the emission from the cathode point is approximated by a rectangular distribution over a 15–20 degrees circular area. The radius of this area is denoted by "a". For small separations, the electrostatic dispersion of the emitted electron beam may be neglected and an equal area of radius "a" at the anode is
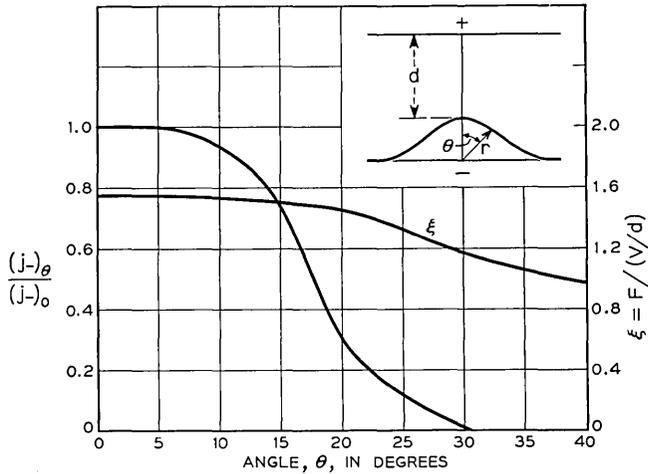

Fig. 4 — Illustration of field and current density distributions for a two-dimensional case: $V/d = 30 \times 10^6$ volts/cm, $\varphi = 4.0$ e.volts and $d = r/2$.

bombarded. A solution was worked out for the resulting temperature rise for points in the anode metal along the axis of the bombarded area:

$$(\Delta^T)_{z,\,t} = \frac{j_- V a}{k} \left[\frac{2}{a} \left(\frac{\alpha t}{\pi} \right)^{1/2} \left(e^{-(z^2/4\alpha t)} - e^{-(a^2+z^2/4\alpha t)} \right) \right.$$

$$\left. - \frac{z}{a} \operatorname{erfc} \frac{z}{(4\alpha t)^{1/2}} + \frac{1}{a} (a^2 + z^2)^{1/2} \operatorname{erfc} \left(\frac{a^2 + z^2}{4\alpha t} \right)^{1/2} \right] \tag{3}$$

At the anode surface, the temperature rise at the center of the bombarded disc is

$$(\Delta^T)_{z=0} = \frac{j_- V a}{k} \left[(1 - e^{-u^2})/\pi^{1/2} u - \operatorname{erf}(u) + 1 \right] \tag{3'}$$

where $u = a/(4\alpha t)^{1/2}$ and the asymptotic value is:

$$(\Delta^T)_{\max.,\,z=0} = j_- V a/k \tag{3''}$$

Numerically, at $j_- = 10^5$ amp/cm^2, $V = 100$ volts, $a = 10^{-4}$ cm and $k = 0.5$ watt/cm °C, the temperature rise is 2,000°C. Equation (3″) may be rewritten in terms of the field $F = \xi V/d$:

$$(\Delta^T)_{\text{max.}} = (j_- F)(d/\xi)(a/k) \qquad (4)$$

By combining this equation with the emission equation (2), j_- is eliminated and for any separation d and cathode work function φ the temperature rise is calculated as a function of the applied field.

In studies of vacuum breakdowns initiated by field emission, it has been repeatedly observed that a sudden increase in emission is obtained by positive ion formation following an evaporation process.[6, 11 to 15] In studies with point cathode and large cathode-anode separations, evaporation has been attributed to the joule heating of the cathode. Also the rupturing of a point on the cathode due to joule heating and electrostatic pulling has been suggested. In these cases, one may assume that vaporization is generally due to a heated cathode rather than heating of the anode. The temperature rise of the latter is usually comparatively small due to the electrostatic dispersion of the electron beam.* At small separations, however, where the dispersion of the electron beam is small, and for the same material for anode and cathode, the anode heating becomes more significant and the anode may evaporate before the cathode. This has been previously observed and reported. For instance,[16] using a rotating mirror camera studies of contacts separated at about a millimeter have rather consistently indicated *anode* evaporation to precede that of the cathode. It will be shown here, by a simple calculation that for small separations, one may generally assume the anode to evaporate before the cathode. For this we will consider an extreme case, where the contact geometry is so chosen that the temperature rise at the cathode is maximum and at the anode is minimum. The cathode emitting point is assumed to be a cylinder of radius a and height ℓ with its base maintained at temperature T_o. The anode is assumed to be a semi-infinite body with a plane disc of the same radius a heated by electron bombardment. The temperature rise at the cathode due to joule heating is given by:

$$(\Delta^T)_- = I^2 \rho \ell^2 / 8.36 \pi^2 k a^4$$

The temperature rise at the anode is obtained from equation (3″)

$$(\Delta^T)_+ = IV/\pi a k$$

---

* At longer gaps where high voltages are required for breakdown, it has been suggested[17] that the high energy electrons bombarding the anode may cause the emission of positive ions and photons which in turn will cause further emission from the cathode.

where $I = j\pi a^2$. The ratio of the two is given by:

$$(\Delta^T)_+/(\Delta^T)_- = 26a(a/\ell)^2 V/I\rho$$

Numerically, for $a = 10^{-4}$ cm, $V = 100$ volts, $I = 10^{-3}$ ampere and $\rho = 10^{-5}$ ohm cm, the ratio of anode to cathode temperatures is 2.6 $\times 10^7 (a/\ell)^2$. For the usual values of a/ℓ commonly obtained on contact surfaces one concludes that the anode heating is more efficient and the anode is the source of vapor for the ion formation necessary for producing a breakdown.

If the mechanism of the breakdown simply involves cathode field emission followed by anode evaporation, one may substitute $T_{max.} = T_b$, the boiling temperature of the anode, and calculate the relation between

TABLE II — ARC INITIATION FIELDS FOR $d = 800\text{Å}$, $k = 0.7$ watt/cm. °C, $a = 10^{-4} - 10^{-3}$ cm, and $T_{max} = 2{,}500°\text{K}$ (boiling) and $1{,}000°\text{K}$ (softening).

	(1)	(2)	(3)	(4)	(5)
(1) φ e. volts	3.0	3.5	4.0	4.5	5.0
(2) F calculated: $10^7 V$/cm T max. = 2500°k	24–21	30–26	36–32	43–38	50–45
(3) F calculated: $10^7 V$/cm $T_{max} = 1000°$k	22–19	27–23	33–29	39–35	45–41
(4) F measured: $10^7 V$/cm			30 (Commercial Palladium)		

contact separation and arc initiation voltage. Table II, line 2, gives calculated values of the arc initiation field for: $d/\xi = 800\text{Å}$,* a range of $\varphi = 3$ to 5 e.volts and $a = 10^{-4}$ and 10^{-3} cms.† In these calculations the effects of electrostatic pulling have not been considered. While the arc initiation field has been found to be below the field necessary for electrostatic pulling of the cold metal one should consider the effect of anode heating and the possible lowering of its tensile strength. For most metals the tensile strength decreases with temperature rise. For instance, for Pd^5 the tensile strength decreases by a factor of 10 when heated from normal temperature to about 1,100°C. Furthermore, most metals have a so-called "softening temperature" which is about $\frac{1}{3}$ to $\frac{1}{5}$ the melting temperature.[19] One may, therefore, expect that heating of the anode only to the softening temperature by electron bombardment should be

* For the small separations used in this study, ξ is estimated at 1.0 to 1.1.
† These values of a correspond to the sizes of surfaces irregularities measured from cross-sectional photomicrographs of the contacts used in our experiment.

sufficient to pull a point electrostatically and increase the field by decreasing the separation d. This corresponds to an increase in the intensification factor ξ of the apparent field V/d. In Table II, line 3, calculations are given of the arc initiation field for $T_{\max.}$ = softening temperature instead of the boiling temperature.

It is of interest to show from the above analysis, that the constant field criterion obtained experimentally for arc initiation is only approximately true and the field for arc initiation should decrease with increasing separation. Equation (4) shows that $j_-F\alpha1/d$. The emission equation (2) may be approximated by: $j_-\alpha F^n$ where n is between 10 and 30 for $\varphi = 3\text{--}5$ e.volts and for fields of the order of 10^7 volts/cm. By eliminating j_- one obtains $F^{n+1}\alpha1/d$ and $V_{ai}\alpha(d)^{1-(1/n+1)}$ or $V_{ai}\alpha d^{0.9-0.97}$.

LIMITING CONDITIONS FOR ARC INITIATION — CONCEPTS OF LIMITING POWER AND ARC INITIATION CURRENT

Replacing j_- in equation (3″) by the total current $I = \pi a^2 j_-$ one obtains:

$$IV = \pi ak\Delta T \tag{5}$$

To produce a breakdown the temperature should rise to the boiling temperature of the metal with or without the effect of electrostatic pulling. Consider the circuit of Fig. 5 (a) consisting of an adjustable voltage source, a contact and a fixed resistance R. Let the applied voltage be V_o and the contact separation is gradually decreased. For each value of field there will be a field emission current I such that the voltage drop across the contact is given by $V_c = V_o - IR$. This relation is shown in Fig. 5 (b) as V_c versus I. For each contact separation there is an emission line as shown schematically in Fig. 5 (b). The point of intersection P of the circuit line and the emission line determines the only possible operating point. If the corresponding power IV is not sufficient to allow a temperature rise $(\Delta T)_b$, equation (5), no breakdown will occur. One can then decrease the contact separation further and the point P will slide along the circuit characteristic. The value of the power IV will increase to a maximum at $V_c = V_o/2$. This maximum power is $V_o^2/4R$. If this is still not sufficient to produce the temperature rise $(\Delta T)_b$ no breakdown can occur and closure without breakdown is obtained. The condition for obtaining a breakdown is, therefore, given by:

$$V_o^2/R \geqq 4\pi ak(\Delta T)_b \tag{6}$$

To verify this rather simple relation, the circuit in Fig. 5 (a) was used. The resistor R is a carbon resistor with the contact metal mounted directly at its end to reduce the effects of local capacities. For each value of R a limiting value of applied voltage V_o was observed below which contact closure occurred without breakdown. The results are plotted in Fig. 6 for Pd contacts. The results are fitted by the parabola

$$(V_o^2/R)_{\text{Crit.}} = 4.2 \text{ watts}$$

Using $(\Delta T)_b = 2{,}200°C$, $k = 0.7$ watt/cm°C, one obtains the value of a from equation (5): $a = 4.2/4\pi k(\Delta T)_b = 2.2 \times 10^{-4}$ cm which is within the range of sizes of surface irregularities measured from cross-sectional photomicrographs of the contacts used. From this one concludes: *To initiate a breakdown there is a minimum circuit power V_o^2/R below which no breakdown will occur on closure. It is determined from the physical properties of the metal and the geometry of the contact surfaces.* The corresponding critical power at the contact $(IV)_{\text{cont.}}$ is, as given by
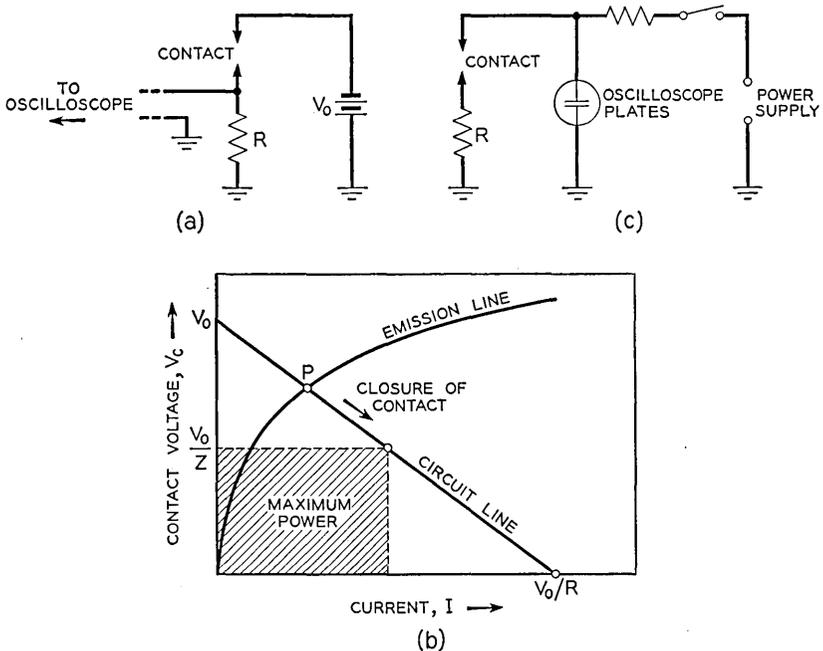

Fig. 5 — (a) Circuit for study of limiting conditions for initiation of a breakdown. (b) Circuit and field emission characteristics. (c) Circuit for observation of unstable breakdowns.
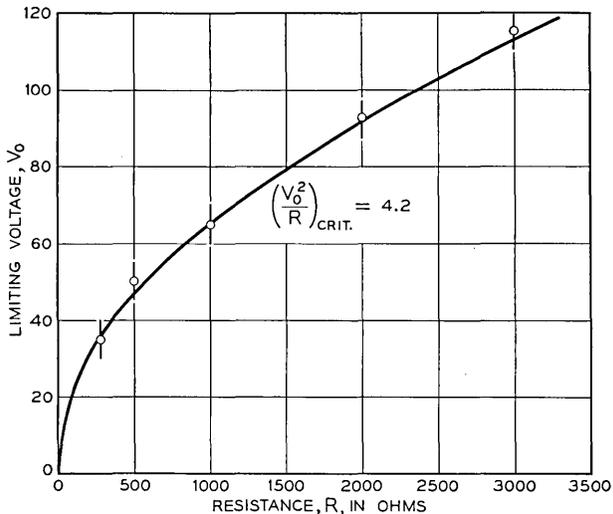
Fig. 6 — Limiting voltages for occurrence of breakdown as a function of circuit resistance for palladium contacts.

equation (5), equal to $V_o^2/4R$ or $\pi a k (\Delta T)_b$. From this one obtains an expression for the breakdown initiation current:

$$(I)_i = \pi a k (\Delta T)_b / V \tag{7}$$

This current is a function of the physical properties of the contact metal, the surface geometry and is inversely proportional to the applied contact voltage. For the Pd contacts experimented with, $(I)_i \cong 1/V$ and for $V = 100$ volts $(I_i) = 0.01$ amp. This is about two orders of magnitude below the minimum arcing current or arc termination current.[2] In other words, *the initiation of a breakdown will not necessarily lead to the establishment of an arc. The latter is only obtained if the circuit is capable of furnish-*
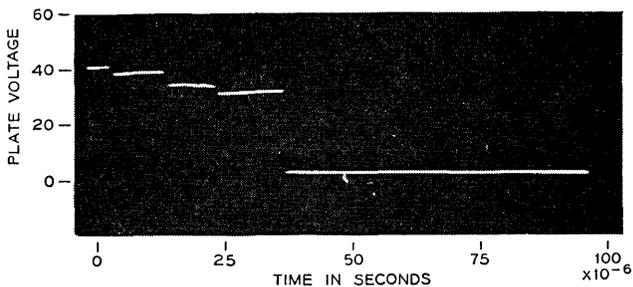

Fig. 7 — Typical transient on closure showing unstable breakdowns obtained with circuit in Fig. 6. $V_0 = 40$ volts, $R = 1000$ ohms and $c = 5 \times 10^{-12}$ farad.

ing a current at least equal to the minimum arcing current. Otherwise, only transient and unstable breakdowns are obtained. These unstable breakdowns have been observed by using the circuit in Fig. 5 (c). The oscilloscope plates were used as the circuit capacitance and the resistance R was so chosen that the circuit could not furnish the minimum arcing current. By slowly closing the contacts transient breakdowns were observed and recorded. Fig. 7 is a typical recording of these breakdowns, followed by a closure.*

TIME LAG IN ARC INITIATION

As discussed above, a breakdown only occurs when the anode temperature is raised by a certain amount $(\Delta T)_c$. The corresponding heating time is called the arc initiation time lag. In this section a relation is derived between time lag and overvoltage and actual observations of time lags are reported. Consider a pair of contacts with a separation d and a constant voltage pulse V is suddenly applied. The temperature rise is given by Equation (3') as a function of time. Let $(V_{ai})_\infty$ be the voltage which gives a temperature rise just equal to the critical value $(\Delta T)_c$ at $t = \infty$. This is the minimum arc initiation voltage which corresponds to an infinite time lag and $(\Delta T)_c = (j\_F)_\infty da/\xi k$. Now let a higher voltage $V = AV_\infty$ be applied where A is slightly greater than 1. This will cause an increase of F_∞ to AF_∞ and of $(j\_F)_\infty$ to $B(j\_F)_\infty$ where B may be obtained from the emission equation (2). Substituting in equation (3'), the temperature rise due to the new voltage is obtained:

$$\Delta T = (\Delta T)_c \, B \, f(u)$$

where $f(u)$ is the function of u given in equation (3'). The breakdown is obtained when $(\Delta T) = (\Delta T)_c$ or $f(u) = 1/B$. Our interest is in values of B very close to 1.0 and $u \ll 1.0$. By expanding $f(u)$ and neglecting higher orders of u:

$$1/B = 1 - u/(\pi)^{1/2}$$

and the time lag is given by:

$$(t)_{lag} = a^2/4\pi\alpha(1 - 1/B)^2 = a^2/4\pi\alpha(B - 1)^2 \tag{8}$$

for $B - 1 \ll 1.0$.

By definitions

$$B(jF) = (jF) + \frac{d(jF)}{dF} \Delta F \quad \text{or} \quad B - 1 = \frac{d(jF)}{dF} \frac{\Delta F}{jF}$$

By multiplying equation (2) by F and differentiating with respect to

* Similar transients have been previously reported by Germer.[18]

F, one obtains:

$$B - 1 = \frac{\Delta F}{F} [3 + 6.83 \times 10^7 \varphi^{3/2} f(y)/F]*$$

Substituting in equation (8) the time lag is obtained as a function of the overvoltage $\Delta V/V$ $(= \Delta F/F)$:

$$(\alpha t/a^2)_{\text{lag}} = \frac{1}{4\pi} \left(\frac{\Delta V}{V}\right)^{-2} \left(3 + \frac{6.83 \times 10^7}{F} \varphi^{3/2} f(y)\right)^{-2} \qquad (9)$$

i.e., the time lag is inversely proportional to the square of the overvoltage. Numerically for $F = 3 \times 10^7$ volts/cm, $\varphi = 4$ e.volts, $a = 5 \times 10^-$ cm and $\alpha = 0.1$ cm^2/sec:

$(t)_{\text{lag}} = 0.65 \times 10^{-9} (\Delta V/V)^{-2}$ and for $\Delta V/V = 0.01$, $(t)_{\text{lag}}$

$$= 6.5 \times 10^{-6} \text{ sec}.$$

These time lags have been observed for palladium and gold contacts.[†] These were obtained by applying a constant voltage pulse, rise time less than 10^{-7} second, across the contacts and recording the voltage at the contacts from a cathode ray oscilloscope. Fig. 8 shows a typical record showing a few voltage pulses applied in succession starting with the low voltage pulse. In this case the time lag was about 15×10^{-6} second. The time lags observed ranged between 20×10^{-6} second and less than 10^{-7} with the majority in the lower end of this range. Similar time lags were also observed by recording the current flow through the contact gap. Fig. 9 is a typical recording of the field emission currents preceding arc initiation. The rise and drop at the left end of the trace are due to the charging and discharging of the oscilloscope circuit following the
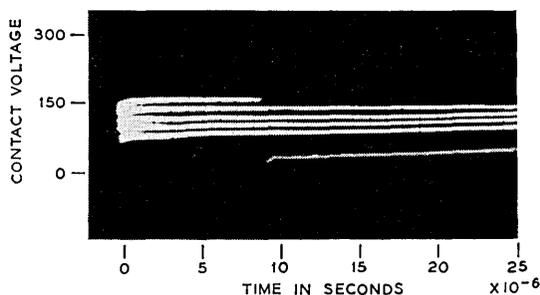

Fig. 8 — Typical recording of the time lag preceding the initiation of the short arc.

* In differentiating, the Nordheim elliptic function $f(y)$ was assumed constant.
† Due to the spread of our data on arc initiation voltage no attempt was made to correlate observed time lags with over-voltage to check Equation 9.
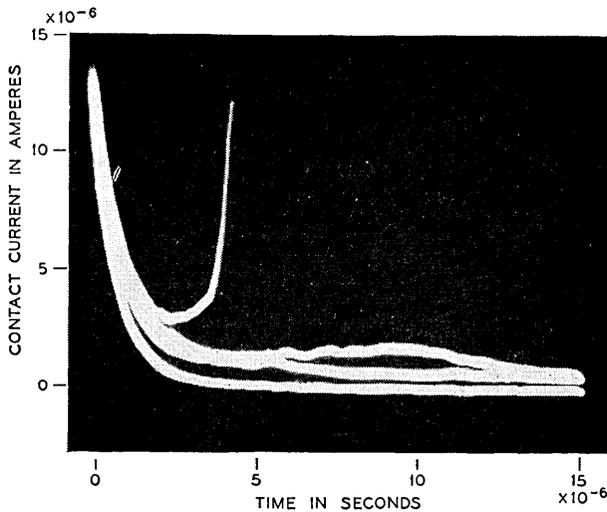
Fig. 9 — Field emission currents preceding arc initiation. Only the top line leads to an arc.

application of the pulse. The lower lines show field emission currents without arc initiation.* The top line shows a transition into an arc.

ACKNOWLEDGMENT

I am indebted to M. G. Reich for assistance with most of the experiments reported here.

BIBLIOGRAPHY

1. L. H. Germer and F. E. Haworth, J. Appl. Phys. 20, pp. 1085–1109 (1949).
2. M. M. Atalla, B.S.T.J., 32, pp. 1493–1506 (1953).
3. P. Kisluik, J. Appl. Phys., 25, p. 897, (1954).
4. G. L. Pearson, Phys. Rev., 56, p. 471 (1939).
5. The Platinum Metals and their Alloys, The International Nickel Co., Inc., (1941).
6. The Physics of Powder Metallurgy, Chap. 8, edited by W. E. Kingston (McGraw-Hill Book Co., N. Y., 1951).
7. C. Herring and J. K. Galt, Phys. Rev., 85, 1060 (1952).
8. A. Sommerfeld and H. Bethe, Handbuch der Physik (Verlag. Julius Springer, Berlin, 1933), 24, No. 2, p. 441.
9. L. W. Nordheim, Proc. Roy. Soc., A121, p. 626 (1928).
10. W. W. Dolan, Phys. Rev., 91, p. 510 (1953).
11. Hull and Burger, Phys. Rev., 21, p. 1121A (1928).
12. Snoddy, Phys. Rev., 37, p. 1678A (1931).
13. Beams, Phys. Rev., 44, p. 803 (1933).
14. Dyke, Trolan, Martin and Barbour, Phys. Rev., 91, p. 1043 (1953).
15. Dolan, Dyke and Trolan, Phys. Rev., 91, p. 1054 (1953).
16. Chiles, J. Appl. Phys., 8, p. 622 (1937).
17. J. G. Trump and R. J. Van De Graaff, J. Appl. Phys., 18, p. 327 (1947).
18. L. H. Germer, J. Appl. Phys., 22, p. 955, (1951).
19. Holm, Electric Contacts, Hugo Gebers Forlag, Stockholm, p. 388 (1946).

* The decay of the emission currents at the end of the trace was due to the decay of the voltage pulse applied which was obtained from an R-C network with long time constant.

Bell System Technical Papers Not Published in this Journal

ALSBERG, D. A.[1]

Transistor Metrolog, Trans. I.R.E., Professional Group in Electron Devices, ED-1, 3, Aug., 1954.

AMES, W. S.,[4] and WEDEL, W. H.[4]

Type-N1 Carrier on Radio and Coaxial Cable, A.I.E.E. Commun. and Electronics, pp. 312–318, Sept., 1954.

ANDERSON, P. W.[1]

Mathematical Model for the Narrowing of Spectral Lines by Exchange or Motion, Phys. Soc. Japan, 9, pp. 316–39, May–June, 1954.

ARNOLD, S. M.[1]

Metal Whiskers — A Factor in Design, Elec. Mfg., 54, pp. 110–114, Nov., 1954, and Electronic Components Symposium, Proc., pp. 38–44, May 4–6, 1954.

BECKER, J. A.[1]

The Life History of Adsorbed Atoms, Ions, and Molecules, Annals of the New York Academy of Sciences, 58, pp. 723–740, Sept. 15, 1954.

BENNETT, W. R.[1]

Sources and Properties of Electrical Noise, Elec. Engr., 73, pp. 1001–1006, Nov., 1954.

BOMMEL, H. E.[1]

Ultrasonic Attenuation in Superconducting Lead, Letter to the Editor, Phys. Rev., p. 220, Oct. 1, 1954.

---

[1] Bell Telephone Laboratories, Inc.
[4] Pacific Telephone and Telegraph Company.

BOND, W. L.[1]

Notes on Solution of Problems in Odd Job Vapor Coating, Opt. Soc. Am. J., **44,** pp. 429–38, June, 1954.

BONNER, A. L.[1]

Servicing Center for Short Haul Carrier Telephone Systems, A.I.E.E. Commun. and Electronics, pp. 388–396, Sept., 1954, and Elec. Engr., **73,** p. 973, Nov., 1954.

BOZORTH, R. M.[1]

Magnetostriction and Crystal Anistropy of Single Crystals of Hexagonal Cobalt, Phys. Rev., **96,** pp. 311–316, Oct. 15, 1954.

BOZORTH, R. M., see Nesbitt, E. A.

BRATTAIN, W. H.,[1] and GARRETT, C. G. B.[1]

Surface Properties of Germanium and Silicon, Annals of the New York Academy of Sciences, **58,** pp. 951–958, Sept. 15, 1954.

BULLARD, W. R.,[5] HAYES, J. B.,[1] and SAUNDERS, H. O.[2]

Electrical Protection of Telephone Systems, A.I.E.E. Commun. and Electronics, pp. 385–388, Sept., 1954.

CALBICK, C. J.[1]

Surface Studies with the Electron Microscope, Annals of the New York Academy of Sciences, **58,** pp. 873–892, Sept. 15, 1954.

CAMPBELL, W. E., see Garn, P. D.

CHANEY, W. G., see Kraus, C. R.

CHARNY, A. L.[6]

Application of Toll Dialing in Pennsylvania, A.I.E.E. Commun. and Electronics, pp. 400–405, Sept., 1954, and Elec. Engr., **73,** p. 990, Nov., 1954.

CHRISTENSEN, H.[1]

Surface Conduction Channel Phenomena in Germanium, Proc. I.R.E., **42,** pp. 1371–1376, Sept., 1954.

---

[1] Bell Telephone Laboratories, Inc.
[2] American Telephone and Telegraph Company.
[5] Ebasco Services, Inc.
[6] Bell Telephone Company of Pennsylvania.

CLARKE, K. B.,[3] and COURAGE, J. W.[3]

Making Small Parts, Electronics, **27,** pp. M15–M22, Oct., 1954.

COE, R. H.,[4] and McCLATCHIE, F. F.[4]

One Approach to a Video SHF Relay System (Digest), Elec. Engr., **73,** p. 921, Oct., 1954.

COE, R. H.,[4] and McCLATCHIE, F. F.[4]

One Approach to a Video SHF Relay System, A.I.E.E. Commun. and Electronics, pp. 323–329, Sept., 1954.

CORENZWIT, E., see Matthias, B. T.

DANIELSON, W. E., see Pierce, J. R.

DEVLIN, G. E., see A. L. Schawlow.

ELMENDORF, C. H.[1]

Component Engineering as a Part of System Design, Electronic Components Symposium, Proc., pp. 8–12, May 4–6, 1954.

FISK, J. B.[1]

Acoustics in Communication, Acous. Soc. Am. J., **26,** pp. 644–645, Sept., 1954.

FLOOD, W. F., see Hrostowski, H. J.

FRANCOIS, E. E., see Law, J. T.

FULLER, C. S.[1]

Solar Battery, The Indicator, Special Convention Number, Sept., 1954.

FULLER, C. S.,[1] and SEVERIENS, J. C.[1]

The Mobility of Lithium Ions in Germanium and Silicon, Phys. Rev., **96,** pp. 21–24, Oct. 1, 1954.

GARN, P. D.,[1] and CAMPBELL, W. E.[6]

A Ball-and-Cup Absolute Microviscometer, Analytical Chemistry. **26,** pp. 1609–13, Oct., 1954.

---

[1] Bell Telephone Laboratories, Inc.
[3] Western Electric Company, Inc.
[4] Pacific Telephone and Telegraph Company.
[6] Bell Telephone Company of Pennsylvania.

GARRETT, C. G. B., see Brattain, W. H.

GEBALLE, T. H.,[1] and MORIN, F. J.[1]

Ionization Energies of Groups III and V Elements in Germanium, Letter to the Editor, Phys. Rev., **95,** pp. 1085–6, Aug. 15, 1954.

GEBALLE, T. H., see Matthias, B. T.

GELLER, S.,[1] and WOOD, E. A.[1]

Crystal Structure of Rhodium Silicide, Acta Cryst., **7,** pp. 441–3, May 20, 1954.

GELLER, S., see Matthias, B. T.

GYORGY, E. M., see Weiss, M. T.

HAGSTRUM, HOMER D.[1]

Auger Ejection of Electrons from Tungsten by Noble Gas Ions, Phys. Rev., **96,** pp. 325–335, Oct., 15, 1954.

HAGSTRUM, HOMER D.[1]

Theory of Auger Ejection of Electrons from Metals by Ions, Phys. Rev., **96,** pp. 336–365, Oct. 15, 1954.

HAYES, J. B., see Bullard, W. R.

HAYNES, J. R., see Hornbeck, J. A.

HORNBECK, J. A.,[1] and HAYNES, J. R.[1]

Decay in Photoconductivity Associated with Deep Electron Traps in p-Type Silicon, Phys. Rev., **94,** pp. 1437–8, June 1, 1954.

HOWARD, R. C., see Savant, C. J.

HROSTOWSKI, H. J.,[1] WHEATLEY, G. H.,[1] and FLOOD, W. F.[1]

Anomalous Optical Behavior of InSb and InAs, Letter to the Editor, Phys. Rev., **95,** pp. 1683–1684, Sept. 15, 1954.

JACOBS, O. B.[1]

Aurora Borealis — A Defense Problem, Wire and Radio Commun., pp. 10–12, Oct., 1954.

---

[1] Bell Telephone Laboratories, Inc.

KLEINMAN, D. A., see Ryder, E. J.

KNAPP, H. M.

Better Designing for Automatic Production — Lower-Cost Telephone Relay Highlights, Automation, 1, pp. 51–56, Sept., 1954.

KOHN, W.,[8] and LUTTINGER, J. M.[9]

Quantum Theory of Cyclotron Resonance in Semiconductors, Letter to the Editor, Phys. Rev., 96, pp. 529–530, Oct. 15, 1954.

KRAUS, C. R.,[6] CHANEY, W. G.,[6] and STEELMAN, A. T.[6]

Radio Transmission of Narrow-Band Mobile Radio Systems at 40 Megacycles, A.I.E.E. Commun. and Electronics, pp. 302–307, Sept., 1954.

LAUBE, O. T.[2]

Technical Considerations Relating to Marine Radiotelephone Communication for Safety, A.I.E.E. Commun. and Electronics, pp. 347–352, Sept., 1954.

LAW, J. T.[1]

A Mechanism for Water Induced Excess Reverse Dark Current on Grown Germanium n-p Junctions, Proc. I.R.E., 42, pp. 1367–1370, Sept., 1954.

LAW, J. T.,[1] and FRANCOIS, E. E.[1]

The Adsorption of Gases and Vapors in Germanium, Annals of the New York Academy of Sciences, 58, pp. 925–936, Sept. 15, 1954.

LEE, C. Y.[1]

Switching Functions on an n-Dimensional Cube, A.I.E.E. Commun. and Electronics, pp. 289–291, Sept., 1954.

LEWIS, H. W., see Schawlow, A. L.

---

[1] Bell Telephone Laboratories, Inc.
[2] American Telephone and Telegraph Company.
[6] Bell Telephone Company of Pennsylvania.
[8] Carnegie Institute of Technology, Pittsburgh, Pa., formerly with Bell Telephone Laboratories.
[9] University of Michigan, Ann Arbor, Mich., formerly with Bell Telephone Laboratories.

LOGAN, R. A.,[1] and SCHWARTZ, M.[1]

  Thermal Effects on Lifetime in Germanium, Phys. Rev., 96, p. 46,
  Oct. 1, 1954.

LUTTINGER, J. M., see Kohn, W.

MAITA, J. P., see Morin, F. J.

MASON, W. P.[1]

  Derivation of Magnetostriction and Anisotropic Energies for Hexago-
  nal, Tetragonal, and Orthorhombic Crystals, Phys. Rev., 96, pp. 302–
  310, Oct. 15, 1954.

MATTHIAS, B. T.,[1] GEBALLE, T. H.,[1] GELLER, S.,[1] and CORENZWIT, E.[1]

  Superconductivity of Nb₃Sn, Phys. Rev., 95, p. 1435, Sept. 15, 1954.

MATTHIAS, B. T., see Schawlow, A. L.

McCLATCHIE, F. F., see Coe, R. H.

MORIN, F. J.,[1] and MAITA, J. P.[1]

  Electrical Properties of Silicon Containing Arsenic and Boron, Phys.
  Rev., 96, pp. 28–35, Oct., 1, 1954.

NESBITT, E. A.,[1] WILLIAMS, H. J.,[1] and BOZORTH, R. M.[1]

  Factors Determining the Permanent Magnet Properties of Single
  Crystals of Fe₂NiAl, J. Appl. Phys., 25, pp. 1014–20, Aug., 1954.

PERRY, A. D.[1]

  Propagation of Electromagnetic Waves in Ferrites, Electronic Com-
  ponents Symposium, Proc., pp. 135–140, May 4–6, 1954.

PETERS, H.[1]

  Hard Rubber, Ind. and Engr. Chem., 46, p. 2112, Oct., 1954.

PIERCE, J. R.[1]

  Social Uses of Science, Am. Scientist, 42, pp. 646–650, Oct., 1954.

PIERCE, J. R.[1]

  Theorem Concerning Noise in Electron Streams, J. Appl. Phys., 25,
  pp. 931–933, Aug., 1954.

[1] Bell Telephone Laboratories, Inc.

PIERCE, J. R.,[1] and DANIELSON, W. E.[1]

Minimum Noise Figure of Traveling-Wave Tubes with Uniform Helices, J. Appl. Phys., **25,** pp. 1163–5, Sept., 1954.

PIERCE, J. R.,[1] and TIEN, P. K.[1]

Coupling of Modes in Helices, Proc. I.R.E., **42,** pp. 1389–1396, Sept., 1954.

PLYM, L. M.[10]

Some Experiences with Cathodic Protection in Chicago, Corrosion, **10,** pp. 213–4, July, 1954.

READ, W. T., JR.[1]

Theory of Dislocations in Germanium, Phil. Mag., **45,** pp. 775–796, Aug., 1954.

REMEIKA, J. P.[1]

New Method for Growing Barium Titanate Single Crystals, Electronic Components Symposium, Proc., pp. 61–62, May 4–6, 1954.

RIGTERINK, M. D.[1]

Ceramic Cores for Deposited Carbon Resistors, Ceramic Age, **63,** pp. 10–13, Jan., 1954.

Ross, I. M., see Ryder, E. J.

RYDER, E. J.,[1] Ross, I. M.,[1] and KLEINMAN, D. A.[1]

Electron Multiplication in Germanium at Low Temperature, Letter to the Editor, Phys. Rev., **95,** pp. 1342–3, Sept., 1954.

SAUNDERS, H. O., see Bullard, W. R.

SAVANT, C. J.,[11] and HOWARD, R. C.[1]

Multiplier for Analog Computers, Electronics, **27,** pp. 144–147, Sept., 1954.

SCHAWLOW, A. L.,[1] MATTHIAS, B. T.,[1] LEWIS, H. W.,[1] and DEVLIN, G. E.[1]

Structure of the Intermediate State in Superconductors, Letter to the Editor, Phys. Rev., **95,** pp. 1344–1345, Sept. 1, 1954.

SCHWARTZ, M., see Logan, R. A.

SEVERIENS, J. C., see Fuller, C. S.

SHARPLESS, W. M.[1]

A Calorimeter for Power Measurements of Millimeter Wavelengths, I.R.E. Trans. P.G.M.T.T., MTT-2, No. 3, p. 45, Sept., 1954.

SMETHURST, J. O.[1]

E-Type Telephone Repeaters, A Digest, Elect. Engr., p. 937, Oct., 1954.

SOUTER, J. C.[3]

Synthetic Resins as Coatings and Castings, Electronic Components Symposium, Proc., pp. 24–29, May 4–6, 1954.

STEELMAN, A. T., see Kraus, C. R.

TIDD, E. D.[1]

Recent Trends in Terminals for Hermetically Sealed Components, Electronic Components Symposium, Proc., pp. 112–6, May 4–6, 1954.

TIEN, P. K.[1]

Focusing of a Long Cylindrical Electron Stream by Means of Periodic Electrostatic Fields, J. Appl. Phys., 25, pp. 1281–1288, Oct., 1954.

TIEN, P. K., see Pierce, J. R.

VOGT, E.,[1] and WANNIER, G. H.[1]

Scattering of Ions by Polarization Forces, Phys. Rev., 95, pp. 1190–1198, Sept. 1, 1954.

WANNIER, G. H., see Vogt, E.

WEDEL, W. H., see Ames, W. S.

Weiss, M. T.,[1] and Gyorgy, E. M.[1]

Low Loss Dielectric Waveguides, Trans. I.R.E., Prof. Group on Microwave Theory and Techniques, MTT-2, No. 3, p. 38, Sept., 1954.

Wheatley, G. H., see Hrostowski, H. J.

Williams, H. J., see Nesbitt, E. A.

Wolff, P. A.[1]

Theory of Electron Multiplication in Silicon and Germanium, Phys. Rev., **95,** p. 1415, Sept. 15, 1954.

Wood, E. A., see Geller, S.

---

[1] Bell Telephone Laboratories, Inc.

Recent Monographs of Bell System Technical Papers Not Published in This Journal*

AHEARN, A. J., see Hannay, N. B.

ARNOLD, S. M.

Metal Whiskers — A Factor in Design, Monograph 2338.

BEACH, A. L., see Guldner, W. G.

BIDDULPH, R.

Short-Term Autocorrelation Analysis and Correlatograms of Spoken Digits, Monograph 2268.

BOZORTH, R. M., see Nesbitt, E. A.

BRATTAIN, W. H., see Garrett, C. G. B.

BULLINGTON, K.

Reflection Coefficients of Irregular Terrain, Monograph 2226.

COOK, J. S., KOMPFNER, R., and SUHL, H.

Nonreciprocal Loss in Traveling-Wave Tubes Using Ferrite Attenuators, Monograph 2235.

ELMENDORF, C. H.

Component Engineering as a Part of System Design, Monograph 2337.

FRANÇOIS, E. E., see Law, J. T.

GARRETT, C. G. B., and BRATTAIN, W. H.

Self-Powered Semiconductor Amplifier, Monograph 2311.

GOHN, G. R., GUERARD, J. P., and HERBERT, G. J.

Mechanical Properties of Some Nickel Silver Alloy Strips, Monograph 2287.

GUERARD, J. P., see Gohn, G. R.

GULDNER, W. G., and BEACH, A. L.

Gasometric Method for Determination of Hydrogen in Carbon, Monograph 2341.

GYORGY, E. M., see Weiss, M. T.

HANNAY, N. B., and AHEARN, A. J.

A Mass Spectrograph for the Analysis of Solids and Mass Spectrographic Analysis of Solids, Monograph 2249.

HERBERT, G. J., see Gohn, G. R.

HERRING, C.
Low-Energy Phonons in Thermal Conduction, Monograph 2258.

KARLIN, J. E., see Munson, W. A.

KELLY, H. P.

Differential Phase and Gain Measurements in Color Television Systems, Monograph 2299.

KING, R. A., and MORGAN, S. P.

Transmission Formulas and Charts for Laminated Coaxial Cables, Monograph 2227.

KOMPFNER, R., see Cook, J. S.

LAW, J. T., and FRANCOIS, E. E.

The Adsorption of Gases and Vapors on Germanium, Monograph 2332.

LEGG, V. E., and OWENS, C. D.

Magnetic Ferrites: New Materials for Modern Applications, Monograph 2224.

MERZ, W. J.

Domain Formation and Domain Wall Motions in Ferroelectric BaTiO Single Crystals, Monograph 2266.

MORGAN, S. P., see King, R. A.

MUNSON, W. A., and KARLIN, J. E.

Measurement of Human Channel Transmission Characteristics, Monograph 2272.

NESBITT, E. A., WILLIAMS, H. J., and BOZORTH, R. M.

Factors Determining the Permanent Magnet Properties of Single Crystals of Fe₂NiAl, Monograph 2269.

NYQUIST, H., RICE, S. O., and RIORDAN, J.

The Distribution of Random Determinants, Monograph 2305.

OWENS, C. D., see Legg, V. E.

PERRY, A. D.

Propagation of Electromagnetic Waves in Ferrites, Monograph 2340.

PIERCE, J. R.

A Theorem Concerning Noise in Electron Streams, Monograph 2228.

RICE, S. O., see Nyquist, H.

RIORDAN, J., see Nyquist, H.

ROBERTSON, S. D.

A Broad-Band Helix Traveling-Wave Amplifier for Millimeter Wavelengths, Monograph 2336.

SCHAWLOW, A. L.

Nuclear Quadrupole Resonances in Solid Bromine and Iodine Compounds, Monograph 2250.

SHARPLESS, W. M.

A Calorimeter for Power Measurements at Millimeter Wavelengths, Monograph 2335.

SUHL, H., see Cook, J. S.

TIDD, E. D.

Recent Trends in Terminals for Hermetically Sealed Components, Monograph 2339.

TIEN, P. K.

Bifilar Helix for Backward-Wave Oscillators, Monograph 2225.

WALKER, A. C.

Hydrothermal Growth of Quartz Crystals, Monograph 2298.

WEISS, M. T., and GYORGY, E. M.

Low Loss Dielectric Waveguides, Monograph 2334.

WILLIAMS, H. J., see Nesbitt, E. A.

WOLFF, P. A.

Theory of Secondary Electron Cascade in Metals, Monograph 2236.

WOLFF, P. A.

Electron Multiplication in Silicon and Germanium, Monograph 2307.

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. He is a member of Sigma Xi, Sigma Pi Sigma, Pi Tau Sigma, the American Physical Society, and an associate member of the A.S.M.E.

WALTER H. BRATTAIN, B.S., Whitman College, 1924; M.A., University of Oregon, 1926; Ph.D., University of Minnesota, 1928. Radio section, Bureau of Standards, 1928–1929. Bell Telephone Laboratories, 1929–. Division of War Research, Columbia University, 1942–43. Visiting lecturer, Harvard University, Fall term of academic year 1952–53. Dr. Brattain, co-inventor with Dr. John Bardeen of the point-contact transistor, has been primarily concerned with the study of semi-conductors at the Laboratories. He has also studied frequency standards, thermionics, magnetometers and infrared. Doctor of Sciences, Portland University, 1952; Stuart Ballantine Medal of the Franklin Institute, 1952. Member of Sigma Xi and Phi Beta Kappa. Fellow of the American Physical Society, the American Association for Advancement of Science, and member of the Franklin Institute.

A. GARDNER FOX, B.S., Massachusetts Institute of Technology, 1934; M.S., Massachusetts Institute of Technology, 1935. Bell Telephone Laboratories, 1936–. Since 1944 Mr. Fox has been concerned with the design of microwave amplifiers and with research on millimeter waves at the Holmdel Radio Laboratory. From 1942 to 1944 he designed radio transmission filters and antennas for a fire-control radar system. Prior to this he did research on waveguides, development work on radar,

and designed mobile and airborne radio transmitters. He is currently in charge of a microwave physics group doing research in the millimeter wavelength range. Senior member of the I.R.E.

C. G. B. GARRETT, B.A., Cambridge University (Trinity College), 1946; M.A., Cambridge, 1950; Ph.D., Cambridge, 1950. Instructor in physics, Harvard University, 1950–52. Bell Telephone Laboratories, 1952–. Dr. Garrett has been engaged in research and exploratory development on semiconductor surfaces. Prior to coming to the Laboratories Dr. Garrett's chief work was in the field of low-temperature physics. He is the author of "Magnetic Cooling" (Harvard University Press, 1954). Senior Scholar of Trinity College, Cambridge, 1945; Twisden Student of Trinity College, 1949. Fellow of Physical Society (London).

FRANZ E. HOHN, B.S., McKendree College, 1936; M.S., University of Illinois, 1937; Ph.D., University of Illinois, 1940. Mr. Hohn is associate professor in the Departments of Electrical Engineering and Mathematics at the University of Illinois. He studied applications of mathematical theory to switching while he was a research guest at the Laboratories from June, 1953, to September, 1954. Ford Foundation faculty fellowship, 1951–52. Member of Sigma Xi, Pi Mu Epsilon, American Mathematical Society, Mathematical Association of America, American Council of Teachers of Mathematics and the American Association of University Professors.

STEWART E. MILLER, University of Wisconsin, 1936–39; B.S. and M.S., Massachusetts Institute of Technology, 1941. Bell Telephone Laboratories, 1941–. Since June 1954, Mr. Miller has been Assistant Director of Radio Research at Holmdel and has been in charge of research on guided-wave systems and associated millimeter and microwave techniques. During World War II, he worked on airborne radar systems. He also worked on coaxial carrier transmissions systems. Mr. Miller holds patents in connection with automatic frequency control, an oscillator control scheme and the D-C amplifier. Member of the I.R.E., Eta Kappa Nu, Tau Beta Pi and Sigma Xi.

L. ROBERT SCHISSLER, B.S. in E.E., Lehigh University, 1953; S.M. in E.E., Massachusetts Institute of Technology, 1954. Mr. Schissler, a graduate student in electrical engineering at Massachusetts Institute of Technology, was concerned with the practical realization of electronic

logic circuits while a Summer (1954) employee at the Laboratories. During the Summer of 1953 he studied the applications of matrix algebra to switching circuits at the Laboratories. Member of Phi Beta Kappa, Pi Mu Epsilon, Eta Kappa Nu and the I.R.E.

ARTHUR UHLIR, JR., B.S., M.S. in Ch.E., Illinois Institute of Technology, 1945, 1948; S.M. and Ph.D. in Physics, University of Chicago, 1950, 1952. Dr. Uhlir has been engaged in many phases of transistor development since joining the Laboratories in 1951, including electrochemical techniques and studies on semiconductors. Since 1952 he has participated in the Laboratories' Communications Development Training program, giving instruction in semiconductors. Member of Gamma Alpha, a graduate scientific fraternity, and president of its Chicago chapter in 1951; American Physical Society, Sigma Xi and the Institute of Radio Engineers.

MAX T. WEISS, B.S. in E.E., College of the City of New York, 1943; M.S. in E.E. and Ph.D. in Physics, Massachusetts Institute of Technology, 1947 and 1951. Mr. Weiss worked for the Radio Corporation of America in Camden, N. J. from 1943 to 1944 and then served in the U. S. Navy, working on underwater mines at the Naval Ordnance Laboratory. From 1946 to 1950 he was connected with the Research Laboratory of Electronics at the Massachusetts Institute of Technology. He joined Bell Telephone Laboratories in 1950 and for one year was concerned with research in dielectric waveguide transmission. Since then he has been chiefly concerned with research in microwave properties and applications of ferrites. Member of the Institute of Radio Engineers, the American Physical Society, Sigma Xi, and Eta Kappa Nu.