RCA REVIEW

a technical journal

Published quarterly by

RCA LABORATORIES

in cooperation with all subsidiaries and divisions of

RADIO CORPORATION OF AMERICA

VOLUME XXVIII | JUNE 1967 | NUMBER 2

CONTENTS

PAGE

A Review of the Effect of Imperfections on the Electrical Breakdown of P-N Junctions 175
H. KRESSEL

A Survey of Radiation Effects in Metal-Insulator-Semiconductor Devices 208
K. H. ZAININGER AND A. G. HOLMES-SIEDLE

Time-Domain Analysis of Multiple Parallel Transmission Lines 241
H. AMEMIYA

Unlock Behavior of the Second Order Phase Locked Loop With and Without Interfering Carriers 277
F. S. KEBLAWI

Landau Damping in Electron Beams 297
M. CAULTON

Sonic Film Memory 317
H. WEINSTEIN, L. ONYSHKEVYCH, K. KARSTAD, AND R. SHAHBENDER

Boron Nitride as a Diffusion Source for Silicon 344
N. GOLDSMITH, J. OLMSTEAD, AND J. SCOTT, JR.

A Programming System for Electron Optical Simulation 351
H. E. KULSRUD

A Collision-Induced Instability in Semiconductor Plasmas 366
B. B. ROBINSON

RCA Technical Papers 373

Authors 376

© 1967 by Radio Corporation of America
All rights reserved

RCA REVIEW is regularly abstracted and indexed by *Abstracts of Photographic Science and Engineering Literature, Applied Science and Technology Index, Bulletin Signalétique des Télécommunications, Chemical Abstracts, Electronic and Radio Engineer, Mathematical Reviews, and Science Abstracts* (I.E.E.-Brit.).

RCA REVIEW

BOARD OF EDITORS

Chairman
H. W. Leverenz
RCA Laboratories

A. A. Barco
RCA Laboratories

E. D. Becken
RCA Communications, Inc.

G. H. Brown
Radio Corporation of America

A. L. Conrad
RCA Service Company

E. W. Engstrom
Radio Corporation of America

A. N. Goldsmith
Honorary Vice President, RCA

J. Hillier
RCA Laboratories

R. S. Holmes
Medical Electronics

E. C. Hughes
Electronic Components and Devices

E. O. Johnson
Electronic Components and Devices

E. A. Laport
Radio Corporation of America

H. R. Lewis
RCA Laboratories

G. F. Maedel
RCA Institutes, Inc.

L. S. Nergaard
RCA Laboratories

H. F. Olson
RCA Laboratories

K. H. Powers
RCA Laboratories

J. A. Rajchman
RCA Laboratories

F. D. Rosi
RCA Laboratories

D. F. Schmit
Radio Corporation of America

L. A. Shotliff
RCA International Division

C. P. Smith
RCA Laboratories

W. M. Webster
RCA Laboratories

Secretary
C. C. Foster
RCA Laboratories

Republication and Translation

Original papers published herein may be referenced or abstracted without further authorization provided proper notation concerning authors and source is included. All rights of republication, including translation into foreign languages, are reserved by RCA Review. Requests for republication and translation privileges should be addressed to *The Manager*.

A REVIEW OF THE EFFECT OF IMPERFECTIONS ON THE ELECTRICAL BREAKDOWN OF P-N JUNCTIONS

By

H. KRESSEL

RCA Laboratories
Princeton, N. J.

Summary—A review is presented of the effect of lattice and junction imperfections on the behavior of reverse-biased p-n junctions. The characteristics of microplasma behavior are discussed. The presentation centers on the properties of Si junctions, since these have been the most extensively studied. It is believed, however, that the basic nature and origin of microplasma behavior are generally similar in other semiconductors, although the effect of individual imperfections may vary. Data comparing the theoretically predicted values of breakdown voltage (obtained from photomultiplication measurements) with the typical observed values for mesa Si and low-voltage GaAs devices suggest that the two are not in very substantial variance despite the presence of microplasmas.

I. INTRODUCTION

AVALANCHE multiplication in p-n junctions has been extensively studied in semiconductors since it was shown by McKay[1] to be the process usually responsible for electrical breakdown of reverse-biased junctions. (Internal field emission is dominant only in very narrow junctions.[2, 3]) Certain anomalies were observed, such as bistable behavior of the current in the breakdown region and slope discontinuities in the *I-V* curves. These were interpreted by McKay as due to nonuniform electrical breakdown of the junction. Subsequent investigations have shown that the reverse *I-V* characteristics of p-n junctions are, in fact, typically dominated by imperfections that limit the breakdown voltage to values lower than the so-called "bulk" values (determined by the carrier ionization rates and junction characteristics).

The localized currents that flow in imperfect junctions have been generally called microplasmas. Nonuniform breakdown has been observed not only in Si but in GaAs,[4-6] Ge,[7] and GaP[8, 9] junctions and appears, therefore, to be a general phenomenon. On the basis of the available experimental evidence in Si, Ge, and GaAs, it appears that the electrical microplasma characteristics are quite similar in these

materials. The role of particular lattice defects may, however, be different in the various semiconductors.

The origin of the microplasma sites has been under extensive investigation in recent years, particularly in Si, and a great deal of progress has been made in isolating some of the factors responsible for nonuniform breakdown. It is possible at this time to fabricate junctions that appear to be free of microplasmas.[10] Relatively defect-free junctions exhibiting high values of photocurrent multiplication have found application as solid-state photomultipliers.[11] Such applications do not generally require completely microplasma-free junctions. The important requirement is that the microplasma breakdown voltage be within a volt or two of the bulk value.

The object of this paper is to review the effect of various junction and lattice imperfections on the reverse-bias *I-V* characteristics. Consideration is limited to bulk imperfections, i.e., imperfections that are not connected with surface effects such as the formation of inversion layers or ionic motion.

II. Characteristics of P-N Junctions Exhibiting Nonuniform Breakdown

Nonuniform p-n junction breakdown is typically detectable by one or more of the following symptoms:

(1) Slope discontinuities in the *I-V* characteristics in the breakdown region. An example is shown in Figure 1 for a reverse-biased GaAs diode.

(2) "Soft" reverse bias *I-V* characteristics of the form $I \propto V^n$, where n is a constant greater than 1 (values between 3 and 6 are typical). An example is shown in Figure 2. These characteristics are not to be confused with those of junctions that break down as a result of an appreciable current component resulting from internal field emission. This effect, however, is typical only for diodes fabricated on heavily doped substrates. In Si, the breakdown voltages observed are of the order of 5 volts or less.

(3) Light emission in discrete spots.

(4) Large amplitude fluctuations of the current in the breakdown region. These fluctuations are particularly pronounced as the junction temperature is reduced.

The absence of the above symptoms is not proof that uniform breakdown is observed. In certain samples, for example, the emission

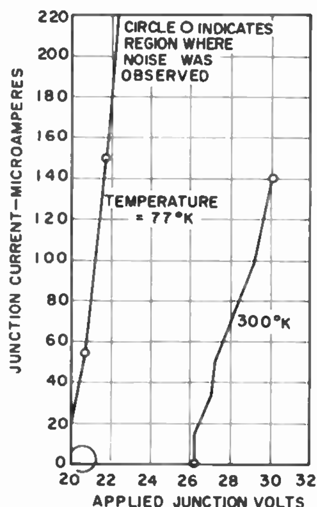


Fig. 1—GaAs diode exhibiting microplasma breakdown (Reference (5)).

of light may not be observed because of reabsorption. Furthermore, not all microplasmas emit light. It is also possible that a smooth *I-V* curve is observed because the current is carried by a single large microplasma; a slope discontinuity may therefore not be readily observed at a convenient current level. A further possibility in diodes fabricated from low-resistivity substrates, is that the slope discontinuities may be difficult to detect because of the relatively low series resistance of the microplasmas. (This is discussed in Section V.) The observation of sharp *I-V* characteristics on a oscilloscope may thus be incorrectly interpreted as representative of bulk breakdown. A number of additional and more refined measurements may, however, be made to determine whether this is in fact the case and, if microplasmas are present, what the bulk breakdown voltage is.

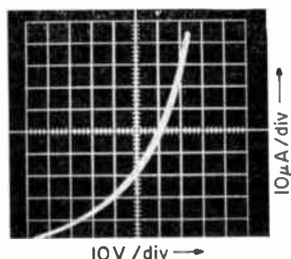


Fig. 2—Typical "soft" reverse *I-V* curve of a silicon diode at room temperature.

In has been shown by Shockley[12] that the degree of junction "patchiness" may be determined from a plot of the current multiplication factor (as determined from a photocurrent multiplication measurement) as a function of the junction voltage;

$$\frac{1}{M} = \frac{n(V_B - V)}{V_B}, \tag{1}$$

where M is the current multiplication factor, V the junction voltage, V_B the bulk breakdown voltage, and n a constant. This expression will hold if the multiplication is uniform over the illuminated region. Deviation will occur if this is not the case, and from these deviations one may estimate the ratio of the microplasma area to the total illuminated area. The value of V_B may also be calculated from these data by extrapolation. Haitz *et al*[13] have investigated the variation of the multiplication coefficient within a given junction by scanning with a small light spot and have shown that Equation (1) holds for values of M up to about 100 in small regions where uniform multiplication may be assumed to take place.

A crude but frequently convenient check on the degree of junction uniformity requires only the use of an oscilloscope. If a junction contains microplasmas that severely reduce the breakdown voltage, then, when the junction is illuminated, the photocurrent variation as a function of voltage will be as shown in Figure 3(a). Very little photocurrent multiplication will be observed before the I-V curve is dominated by the dark current carried by the microplasmas. If, on the other hand, the microplasma breakdown voltage is reasonably close to the bulk breakdown voltage (or the junction breakdown is uniform), then appreciable multiplication will be observed, as shown in Figure 3(b). This simple test is based, of course, on the assumption that the total microplasma region (or regions) is small compared to the total junction area. This assumption is generally reasonable, since microplasmas typically tend to be of the order of a few microns in diameter.

It should be noted that great care is required in interpreting noise characteristics in the breakdown region of very small diodes, since the distinction between small diodes and microplasmas tends to become blurred. The noise characteristics of such diodes have been studied in detail.[14] The absence of noise pulses is not necessarily indicative of microplasma-free junctions. As shown by Kikuchi[15] and Kikuchi and Tachikawa,[16] at least four types of microplasma sites appear to exist in Si junctions (Table I), some of which do not exhibit the typical microplasma noise behavior.

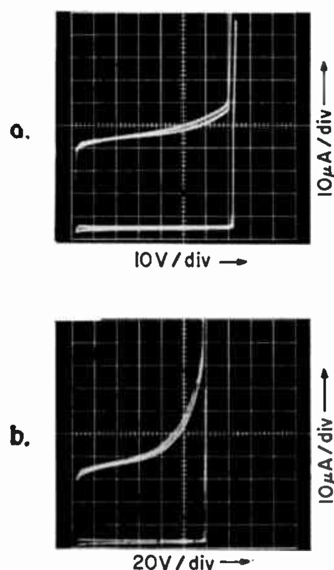


Fig. 3—(a) Oscilloscope trace of silicon junction containing a microplasma that breaks down at 70 volts (in the dark and under illumination). Note the very small degree of photocurrent multiplication. (b) Oscilloscope trace of more uniform silicon device fabricated on the same substrate as the diode shown in (a). Note the change in the voltage scale and the shape of the photocurrent versus voltage curve.

Having determined that the breakdown voltage is imperfection-limited, the question remains whether the surface of the device is at fault or whether a bulk defect is responsible. In the case of mesa diodes, for example, the question may frequently be settled by a repeated etching and surface-cleaning procedure that should improve the device if the surface is at fault. In planar oxide-passivated junctions, the diagnosis is more difficult without permanent changes in the device, such as removal of the oxide. An indication of ionic motion on the junction surface may, however, be obtained from "walk out"

Table I—Classification of Weak Spots in Junction (after Kikuchi, Ref. (15))

Type	Noise Pulses	Light Emission	Local Multiplication	*I-V* Character
I	yes	yes	no	hard
II	yes	no	no	hard
III	no	yes	yes	soft
IV	no	yes	no	soft

effects of the breakdown voltage, i.e., a significant increase of breakdown voltage under bias (where the possibility of junction heating is eliminated). Soft characteristics are also frequently associated with surface effects, although, as discussed later, metallic precipitates result in a similar effect. Excessive values of saturation current combined with a relatively "hard" curve is frequently indicative of the presence of an inversion layer.

Electron-beam scanning techniques have recently been applied to study defects in p-n junctions.[17-19] The method consists of scanning the junction surface with a high-voltage electron beam a few microns

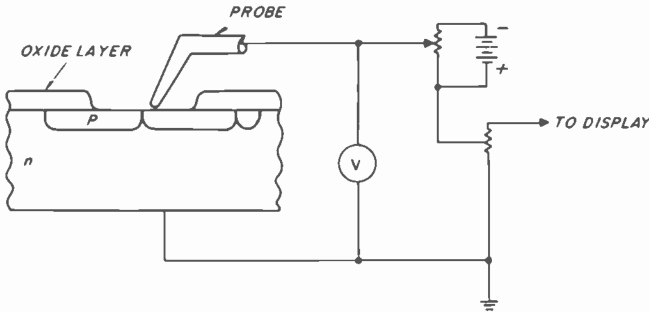


Fig. 4—Schematic of scanning electron-beam microscope apparatus (Reference (17)).

in diameter (or smaller if possible), using an apparatus of the type shown in Figure 4.[17] The electron–hole pairs that are generated by the beam produce a current that is used to modulate the beam current of a cathode-ray display tube. Variations in the collected current arise from local variations in the recombination rate or electric field in the space-charge region. If the beam hits a high-field microplasma region, the generated current will be enhanced by avalanche multiplication. Such a microplasma site should, therefore, appear as a region of above-average brightness on the cathode-ray-tube display. This effect was observed by Gaylord[17] and correlated with localized light emission. A region of high recombination rate, on the other hand, will appear as a spot of below-average brightness. This technique has been used to detect high-recombination-rate regions in mechanically damaged junctions (see Section VII).

III. Photon Emission in Back-Biased Junctions

The subject of light emission from back-biased junctions is closely related to the general area of electroluminescence in semiconductors.

The subject has been reviewed,[20-23] and the following discussion is therefore very brief.

The emission of light was initially reported by Newman in reverse-biased Si junctions.[24] It was shown by Chynoweth and Pearson[25] that the spotty emission observed could be correlated with dislocations, although only a small fraction of the dislocation sites emitted light. Uniform light emission is observed in small-area microplasma-free junctions.[14] The emission of light at various microplasma sites has been investigated by, among others, Kikuchi.[15] As shown in Table I, not all microplasma sites give rise to the emission of light, and the detailed correlation is not well understood. Emission of light at some microplasma sites has also been observed in Ge,[16] GaAs,[21] and GaP.[21]

The photon emission is frequently reabsorbed and, hence, difficult to detect, since the photon energy distribution extends to values well above the band-gap energy. The distinguishing feature of this emission in Si and Ge is that the spectrum is generally broad as compared to the emission from forward-biased junctions. In the case of Si, for example, the emission peaks at about 1.1 eV but the tail extends upward up to about 3 eV.[26] This high-energy distribution may be understood in terms of the recombination of holes with hot electrons having energies up to the ionization energy.

The low-energy tail of the photon distribution may be due to recombination through centers in the forbidden band. A Bremsstrahlung mechanism has also been suggested.[27] In the case of GaAs, however, the emission spectrum is comparable in width to the one observed when the same diodes are forward biased,[23] although the peak energy is shifted to a somewhat lower energy. The difference between the emission in GaAs and in the "indirect" materials, Si and Ge, may be explained by the fact that the recombination of energetic electrons and holes in GaAs is considerably less probable than the recombination of the relatively low-energy electrons located at the $\bar{k} = 0$ minimum of the conduction band. A fraction of the hot electrons may, in fact, be located in the "heavy mass" band. Radiative transition from the bottom of this band is therefore relatively improbable.

IV. Electrical Characteristics of Microplasmas

A number of detailed studies have been made in recent years of the electrical characteristics of diodes exhibiting breakdown through a single or a very small number of microplasmas in order to gain further understanding of the nature of the microplasma sites. Particular attention has been paid to the noisy behavior of the microplasma current flow that is frequently observed and that was originally

described by McKay.[1] At a given voltage, the current pulses have constant amplitudes with rise and decay times of the order of 10^{-8} sec. The pulse intervals and lengths have random values. As the applied voltage is increased, the pulse amplitudes increase to some extent, but the most significant effect is a lengthening of the pulses and an increase in the pulse repetition rate. Further increases in voltage result in the appearance of new sets of higher amplitude current pulses superimposed on the earlier set of pulses. These now give the effect of a continuous background current. McKay explained this behavior by

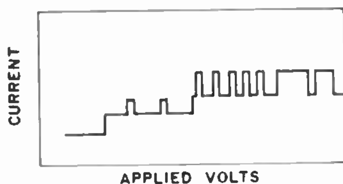


Fig. 5—Schematic of typical microplasma pulse characteristics as a function of applied voltage.

suggesting that one observes the successive contribution of a number of microplasmas, each of which is triggered at some particular voltage. A schematic sketch of the observed behavior is shown in Figure 5.

It is evidently not possible to measure individual microplasma parameters in very soft junctions of the type shown in Figure 2, since the observed behavior may be due to a multitude of microplasma sites. In junctions with *I-V* characteristics in the breakdown region like those shown in Figure 1, however, the contributions of the individual sites to the junction current may be measured. The electrical effects of microplasmas in these junctions have been described in terms of four parameters: breakdown voltage V_B, series resistance R_s, turn-on probability P_{on}, and turn-off probability P_{off}. The *breakdown voltage* is defined as the voltage at which appreciable current flow occurs through a microplasma region. The *series resistance* is obtained from the slope of the linear *I-V* curve in the breakdown region. Typical values are of the order of 1-30 kilohms. The *turn-on* and *turn-off* probabilities are defined as follows. Let the average (d-c) current be $<i>$, the instantaneous current be i (as observed on an oscilloscope), and the number of current pulses per unit time be N. Then,

$$P_{on} = \frac{N}{1 - \dfrac{<i>}{i}}, \tag{2}$$

and

$$P_{off} = \frac{N}{\dfrac{<i>}{i}}, \qquad (3)$$

The variation of P_{on} and P_{off} as a function of temperature and hydrostatic pressure has been studied in Si by Meada and Suzuki.[28] The variation of P_{on} with applied voltage and temperature was found to be of the form

$$P_{on} = A \exp \left\{ B(V - V_B) - \frac{E}{kT} \right\}, \qquad (4)$$

where A is a constant and E varies from 0.14 to 0.23 eV, depending on the sample. B decreases linearly with decreasing temperature and its variation in a linearly graded junction is given by

$$\frac{1}{B}\frac{dB}{dT} = -6 \times 10^{-3}\, C^{-1}. \qquad (5)$$

P_{off} was found to be independent of temperature.

The microplasma noise spectrum was studied by Champlin[29] and Burgess.[30] The measured spectrum extends into the microwave region.[31]

While a detailed explanation of all of the observed microplasma phenomena is not possible at this time, some of the qualitative concepts are believed to be well founded. The basic model is that the electric field in some part of the space-charge region is enhanced over the average value. Since the ionization rate $\alpha(\epsilon)$ is usually an exponential function of the electric field, appreciable electron multiplication may occur locally but not in other areas of the space-charge region where the average field is still low. It is also possible, in some cases, that internal field emission occurs in the high-field regions. If the length of the high-field region l and the field ϵ are such that $\alpha(\epsilon)l \cong 1$, then the local current-multiplication factor will approach infinity. In GaAs or Si, for example, a field of about 7×10^5 volts/cm will correspond to a value of $\alpha \cong 10^5$. Hence, the critical length of the high-field region may be as small as 0.1 μ. A carrier that enters this region has a certain probability of initiating a multiplication process. This probability will depend on the electric field in the region and on its extent. Such initiating carriers may consist of either thermally or optically

generated carriers that reach the vicinity of the high-field region. An experiment performed by Maeda and Suzuki[28] showed, as expected, that P_{on} could be increased by illuminating the diode. Also of interest in this connection are experiments[32, 33] that showed that microplasmas interact among themselves through optical coupling—the light emitted at one microplasma could affect the pulsing behavior of other microplasmas presumably through carrier generation as explained earlier.

As discussed by McIntyre,[34] it is not possible to explain the variation of P_{on} with applied voltage in an unilluminated diode by simply assuming that the thermally generated current is uniformly distributed through the junction and that each microplasma receives a share of it. It appears that the currents flowing into the microplasma sites are considerably larger than expected on the basis of the above assumption. Two possibilities were suggested to explain this effect: (1) carriers are generated locally as a result of internal field emission and (2) minority carrier channeling into the microplasma site occurs. It was subsequently shown[35] that the carrier-generation rate is, in fact, influenced by internal field emission and, furthermore, that carriers may be trapped during the conduction periods of the microplasma and subsequently released, thus triggering a new current burst.

The basis of the turn-off probability was explained[34] by assuming that a microplasma turns off when, as a result of statistical fluctuation, the carried density in its vicinity happens to be zero. P_{off} therefore increases with decreasing current in the on state, but is essentially temperature independent.

The microplasma series resistance is believed to consist of two components[35]—a spreading resistance, R_s, and a space-charge resistance, R_{sc}. R_s is the resistance due to current flow outside the space-charge region. If the microplasma area is small compared to the pellet thickness, as is usually the case, this part of the resistance may be estimated by assuming that the carriers are generated in a region of diameter d and spread into a semi-infinite semiconductor of resistivity ρ:

$$R_s \cong \frac{\rho}{2d}. \tag{6}$$

The space-charge component R_{sc} is due to the resistance to current flow through the space-charge region, and is of a more complex form. McIntyre obtained expressions for graded and abrupt junctions.

Note that R_s will decrease with decreasing resistivity of the more lightly doped side of the junction. As a result, although microplasmas

may be present, their effect on the *I-V* characteristics may not be readily observed in diodes made from heavily doped material, because the slope of the *I-V* curve in the breakdown region will be relatively steep.

V. Origin of Microplasma Sites

Microplasma sites have been shown to result from the presence of certain lattice imperfections in the space-charge region of the p-n junction.

Dislocations

Several studies have been reported attempting to correlate microplasma sites and dislocations following early experiments[25] that related the spotty light-emission pattern in Si junctions to dislocation sites. It was found that only a small fraction of the dislocation sites (as revealed by etch pits) show light emission. Stair-rod dislocations in epitaxial Si films have also been correlated with soft breakdown behavior and light emission.[36] The effect of other types of defects in Si epitaxial films on breakdown was described by Lawrence and Tucker.[37] Also of interest is the result obtained when a nail-head bond (made at 320°C) was removed in a given region and the region of the bond was examined.[38] Light emission was observed in that region, particularly at the periphery of the bond, showing that some microplasmas may be formed in this manner.

An experiment to determine the effect of dislocations in the starting material was described by Prussin.[39] Soft *I-V* characteristics were found to correlate with the presence of grain boundaries in the vicinity of the junction. There are reasons to believe that clusters of dislocations (such as exist at grain boundaries) rather than isolated dislocations are particularly harmful. In a recent experiment,[40] a study was made of the breakdown characteristics of p^+-n and n^+-p Si diodes containing individual dislocations that were introduced by hot-pressing either before or after junction formation by diffusion. The dislocation densities, as determined by a dislocation etch, were as high as 10^5 cm^{-2}. While a reduction in yield of good diodes was observed, the breakdown voltages of a large fraction of the diodes containing dislocations were still comparable to those of the undeformed control wafers. The degraded diodes typically exhibited *I-V* characteristics of the type shown in Figure 1.

It has been suggested[25] that the effect of dislocations on breakdown (apart from the effect of associated clustered impurities) could be explained by the band-gap energy variation in the dislocation region.

This change in band gap results because part of the crystal in the vicinity of an edge dislocation is compressed and part is expanded. It was estimated that in the compressed region of the crystal, the change could be as large as 0.4 eV. This change would have two effects (assuming that the edge dislocation axis traverses the space-charge region). (1) The electron energy required for impact ionization may be reduced, hence the ionization rate would be increased. (2) Channeling of carriers into the dislocation region is possible. The possibility of appreciable channeling has however been questioned[12] on the grounds that the hot-electron energy is sufficiently large that the electrons are not likely to be significantly affected by the lowered potential in the dislocation region. The first effect discussed above is nevertheless possible.

Dislocations in the starting material may affect electrical breakdown indirectly through enhanced diffusion along their lengths. Spikes in the diffusion front may thus be formed giving rise to possible microplasma sites (see Section VI). This effect may, however, not be particularly important for isolated dislocations.[41]

Since dislocations have been shown to be introduced in the process of diffusion,[42] it does not appear possible for junctions formed in this manner to be completely free of dislocations. Since diffused diodes have been fabricated that do appear to exhibit reasonably uniform breakdown, however, these dislocations may not seriously affect the breakdown behavior in the absence of metallic precipitates.

A serious problem appears to arise from the formation of individual dislocations and dislocation clusters during high-temperature heat treatment in regions that have been previously mechanically damaged. It has been shown that dislocations are formed in Si during oxidation of such regions.[43] Using x-ray transmission diffraction microscopy to identify dislocations, Fairfield and Schwuttke[44] have correlated the formation of microplasma sites in the course of device fabrication with dislocations nucleated at scratches or other regons where the silicon surface was previously mechanically damaged. Tweezer marks and even the use of cotton swabs were suggested as sufficient to nucleate microplasma sites. As is the case with many experiments of this nature, the effect on breakdown may be partially attributed to metallic precipitates at the dislocations, rather than by the dislocations themselves. The role of precipitates is discussed further below.

Nonuniform Impurity Distribution

The role of "pipes" in Si junctions, which arise from the accidental deposition of specks of dust containing impurities like phosphorus, has

been investigated by Shockley and Goetzberger.[41] These pipes may consist, for example, of highly doped n-type regions and hence may result in localized values of the breakdown voltage considerably below the average value for the junction.

It has been suggested[12] that even in defect-free junctions, statistical fluctuations in the impurity distribution are probable and, hence, truly uniform breakdown of large-area devices would be difficult to achieve. The calculated fluctuations in breakdown voltage are, however, rather small (appreciably below one volt in junctions where impurity compensation is negligible). The fluctuations are more significant in diffused junctions, where impurity compensation is, of course, significant.

Precipitates in the Space-Charge Region

A great deal of evidence that has accumulated recently strongly suggests that inclusions in the space-charge region are among the most important factors in the p-n junction degradation. The theoretical aspects of the problem were briefly discussed by Shockley.[12]

The role of either conducting or insulating inclusions in locally increasing the electric field in the space-charge region may be qualitatively understood by considering a spherical inclusion having a dielectric constant different from the medium in which it is immersed.[12] For example, in the case of a spherical precipitate such as SiO_2 in Si, it may readily be shown that the maximum field at the surface of the sphere is approximately 1.5 times larger than the average field in the surrounding medium.

Goetzberger and Shockley[45] investigated the effect of precipitated metallic impurities on the reverse current and breakdown voltage in Si junctions. n+ -p junctions were prepared by diffusing phosphorus into 0.04 ohm-cm p-type Si. Nitrates of Cu, Fe, Mn, and Au were applied to the wafers which were then diffused for 1 hr at 1000° C in a hydrogen atmosphere and cooled to room temperature in six minutes. Such a treatment is expected to produce precipitation of the excess metals, since their solubility at 1000°C is considerably larger than at room temperature. The result of this experiment indicated that the reverse diode *I-V* curve became "soft". A similar result may follow from normal heat treatment for long times at high temperatures, presumably because of the indiffusion of Cu, Ni, and Au. The diffusion coefficients of these metals are orders of magnitude larger than those of phosphorus or boron.[46]

Precipitation nuclei may be formed in two ways: (1) homogeneous nucleation, i.e., aggregation of metallic atoms, and (2) heterogeneous

nucleation, i.e., precipitation at existing sites such as dislocations, voids, or existing precipitates. Homogeneous nucleation usually requires a very large supersaturation, while heterogeneous nucleation does not. Since many possible nucleation sites are usually present, heterogeneous nucleation is more likely to occur.

Dash[47] has shown that copper in Si tends to precipitate at dislocation sites. Schwuttke[48] demonstrated that copper precipitates also tend to form in regions of high oxygen concentration, possibly resulting in the formation of Cu-O complexes. These precipitates can take the form of either platelets that lie on the (100) planes or relatively large needles (20-34 μ in some cases) oriented along {111}.

Electron-microscope observations of copper segregates were made by Lawrence[49] in an attempt to correlate these directly with the observed breakdown behavior of p-n junctions. Two types of precipitates were identified. The first consists of large hexagonal platelets formed at sessile dislocations. These platelets are typically 10 μ wide and a fraction of a micron thick. Their position was correlated with local light emission. The second type consists of smaller dispersed segregates, possibly formed at lattice voids near the surface. No light emission was observed from these. Figure 6 shows how these segregates affect the *I-V* characteristics. The curve for junction "A" (which contained a single large platelet) shows the worst characteristics. The effect of the loosely packed segregates (curve for junction "C") was relatively less detrimental, but still resulted in a 70-fold increase in the current over that of a copper-free junction. A study has also been reported relating the presence of metallic precipitates to the electrical characteristics of controller rectifiers.[50]

An important factor to consider is that the solubility of copper increases with the impurity doping level.[51] Therefore, the copper concentration in the vicinity of the junction is highest during heat treatment at high temperatures, since that region tends to be more heavily doped than the bulk of the crystal. Enhanced copper concentration in the diffused region has been graphically shown by introducing radioactive copper at 600-700°C into boron-diffused planar junctions. Radioautograms showed very clearly that the copper concentration was enhanced in the areas covered by the junctions.[52]

While their electrical effects on junctions have not been directly examined, the formation of SiO_2 inclusions was studied by Kaiser.[53] Large clusters (0.1 μ) were estimated to have been formed after heat treatment of silicon with initial oxygen concentrations of about 10^{18} cm^{-3}. The density of such clusters was estimated to be about 10^{10} cm^{-3}.

The presence of other types of precipitates has been shown to affect the electrical behavior of junctions. For example, Sato[54] has shown that the presence of carbon blocks in the reaction tubes used for preparing epitaxial silicon wafers could result in the formation of carbon precipitates that give rise to microplasma behavior in junctions made from this material.

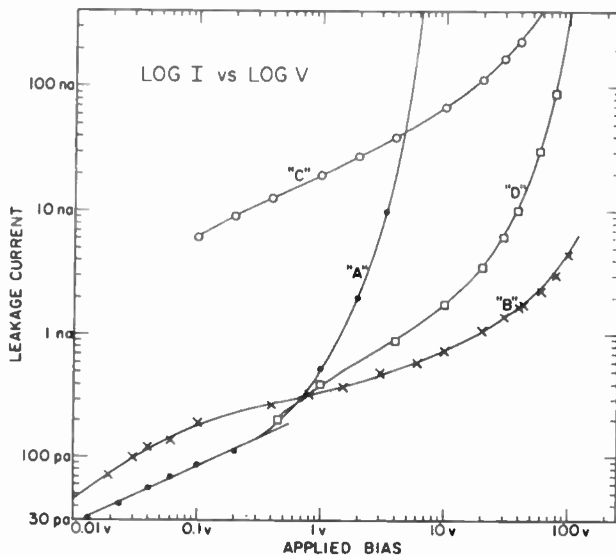


Fig. 6—Leakage current as a function of reverse bias in a planar junction containing metallic precipitates. The origin of the various curves is explained in the text (Reference (49)).

Among other metallic precipitates of interest are those of nickel, a metal which is frequently used in device processing. The precipitation behavior has been studied by Yoshida and Furusho.[55] As expected from the known behavior of copper, nickel precipitates should also be formed near dislocations and may be observed with infrared microscopy.

Gold is frequently introduced into many devices as a minority carrier lifetime "killer" and, as a result, great care is required to prevent the degradation of the reverse *I-V* characteristics. Rapid quenching from high temperature is required in order to minimize the formation of precipitates.

Effect of Radiation Damage

Effects on the reverse-bias characteristics of p-n junctions are likely to arise from both changes in the surface properties of the junction

(particularly in planar devices) and changes in the minority carrier lifetime and hence saturation current. While the forward characteristics have been extensively studied,[56,57] little information is available concerning the effect of radiation damage on breakdown. It is possible that the formation of defect clusters, for example, as a result of heavy particle bombardment, will result in the formation of microplasma sites.

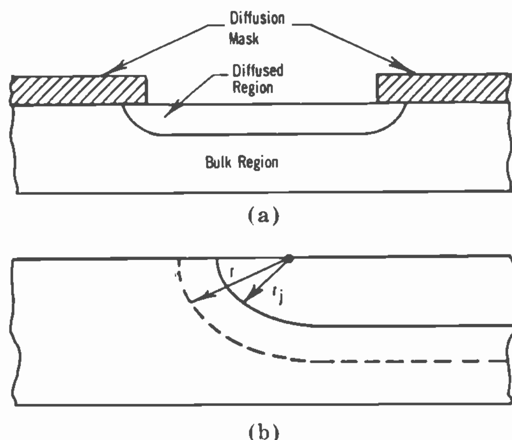


Fig. 7—(a) Cross section of planar diffused junction. (b) Relationship between the diffusion depth and the radius of curvature at the periphery.

VI. Structural Imperfections In P-N Junctions

In addition to lattice imperfections, the breakdown voltage may be limited by certain structural faults in the junction. The best understood of these are the following.

Junction Periphery in Planar Diodes

A common observation is that the breakdown voltage of planar diodes is generally lower than that of mesa diodes fabricated on material of the same resistivity. Figure 7 shows the geometry of a planar junction; as can be seen, the radius of curvature at the junction periphery depends on the diffusion depth. Hilibrand[58] has solved the Poisson equation in cylindrical coordinates to determine the relationship between the field in the junction, the applied voltage, and the radius of curvature. The breakdown voltage was computed[59] using an empirical relationship between α and ϵ of the form $\alpha = 1.95 \times 10^{-24} \epsilon^5$. The resultant plot of breakdown voltages as a function of depth of p+ -n junctions for various values of n-type crystal resistivity is shown in

Figure 8. The arrows on the figure point to the breakdown voltage for a uniform junction. The edge effect diminishes in importance with increasing carrier concentration, because the cylindrical geometry is no longer of significance when the space-charge region width becomes smaller than the radius of curvature. This subject has recently been examined in greater detail.[60]

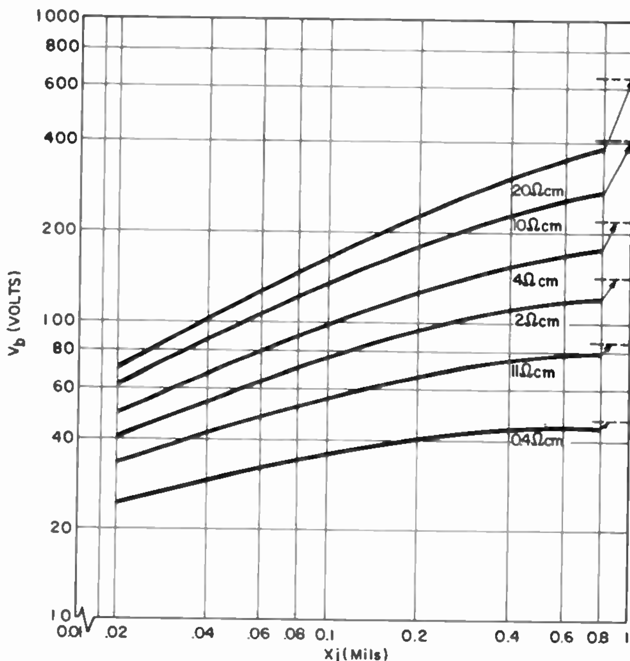


Fig. 8—Breakdown voltage as a function of junction depth for diffused p+-n planar junctions for various values of substrate doping (Reference (58)).

Spikes in the Diffusion Front

Irregularities in the diffusion front of a diffused junction perturb the plane parallel geometry of the space-charge region. This type of defect may give rise to a region of enhanced field as compared to the rest of the junction. One may estimate the effect of such a spike by assuming that it consists of a small hemisphere of radius r_o, as shown in Figure 9. The problem of avalanche breakdown in hemispherical p-n junctions has been treated by Shields.[61] As in the case of the edge of a planar junction, the hemispherical geometry is significant only for radius values that are small compared to the maximum space-charge-region width.

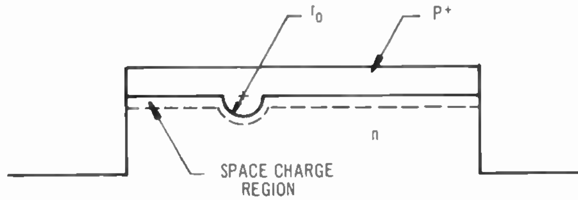


Fig. 9—Cross section of junction containing a spike having a radius of curvature r_0.

Figure 10 shows the critical radius of curvature for various values of the carrier concentration of the lightly doped side of an abrupt Si junction. The effect of the hemispherical geometry may be neglected for values of r_0 greater than the critical values shown. Figure 11 shows the expected breakdown voltage when r_0 is less than the critical value for various values of the impurity concentration.

These irregularities may be caused by the presence of grain boundaries or dislocation clusters that enhance the diffusion coefficient of impurities. In the case of Si, it was found that the diffusion coefficient of phosphorus is enhanced to a greater extent than that of boron. The effect of individual dislocations is believed to be generally unimportant. However, in the case of transistor structures with very thin base regions, the slight junction irregularity due to enhanced diffusion along individual dislocations may result in emitter-to-collector shorts.[41]

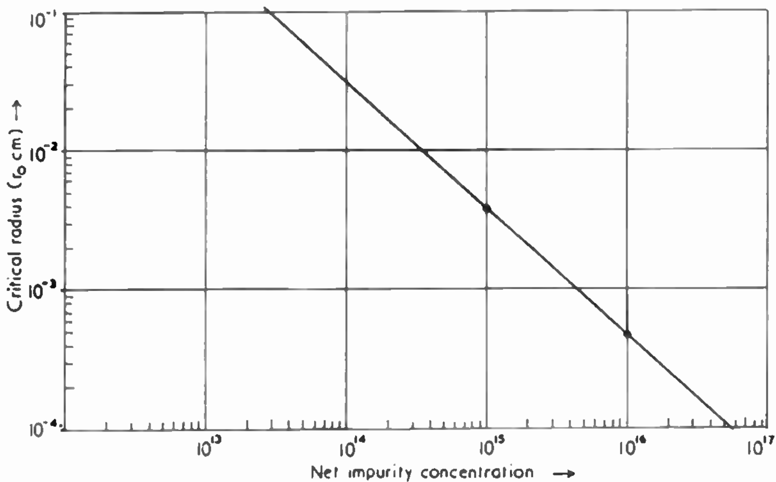


Fig. 10—Critical radius of curvature r_0 as a function of the carrier concentration of the lightly doped side of an abrupt silicon junction (Reference (61)).

VII. Effect of Mechanical Damage

The application of large uniaxial mechanical stresses of the order of 10^{10} dynes/cm^2 to p-n junctions may result in permanent degradation of the reverse *I-V* characteristics—the curve usually becomes "soft." The stress required for serious degradation is difficult to predict, since it may depend not only on the semiconductor but on the tool used to apply the stress and the depth of the junction below the surface. It does appear, however, that the load that a junction may support when the load is applied over a small area (of the order of a few microns in diameter) is considerably in excess of the fracture stress in bulk specimens where the stress would be applied over large areas.

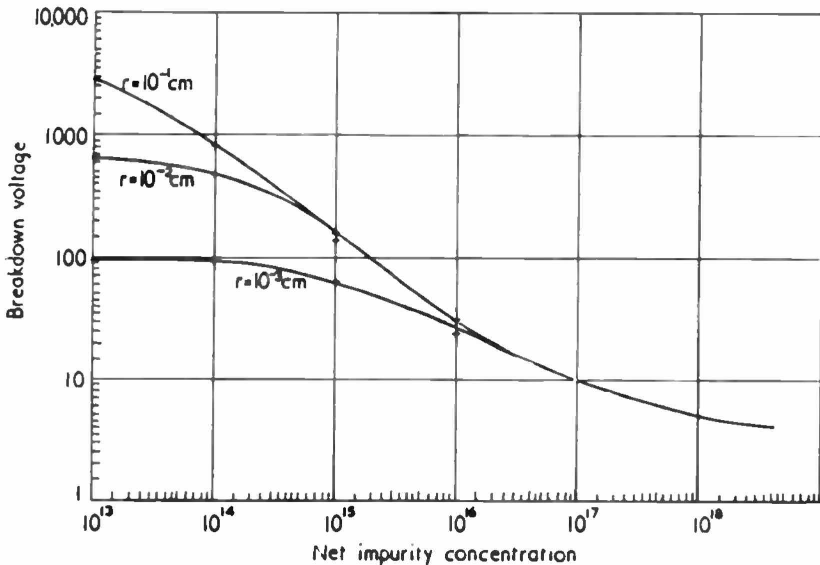


Fig. 11—Breakdown voltage of an abrupt hemispherical silicon diode as a function of the impurity concentration of the lightly doped side of the junction for several values of the radius of curvature (Reference (61)).

Mechanical damage in p-n junctions may be readily detected by increase of the saturation current, which is presumably due to the introduction of recombination centers in the space-charge region and a consequent reduction in the minority-carrier lifetime. Scanning-electron microscope observations support this assumption.[62] In the experiments of Kressel and Elsea,[62] it was found that the excess saturation current began to decrease even after annealing at room temperature.

As shown in Figure 12, the saturation current decreased by several orders of magnitude in the course of isochronal annealing up to 410°C and recovers to a value that is not too far above its original value in the undamaged device. Figure 13 shows that the breakdown voltage of the damaged device may also recover to very nearly its original value

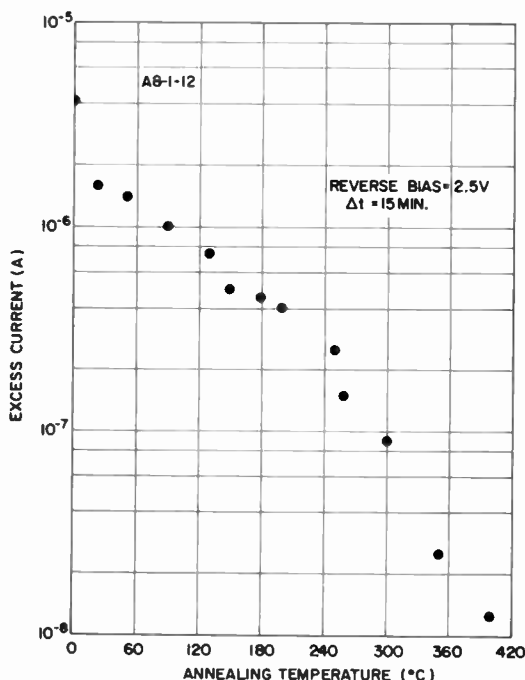


Fig. 12—Excess reverse-bias current (− 2.5 volts) of a mechanically damaged silicon diode as a function of annealing temperature (Reference (62)).

following annealing at the relatively low temperature of 300°C. This temperature is believed to be too low for the removal of dislocations. The recovery of the saturation current and the breakdown voltage at such low temperatures suggests that relatively light mechanical damage may result in the formation of point defects, such as vacancies and vacancy clusters, in addition to dislocations that act as recombination centers. Since the point-defect migration energies are of the order of only one eV,[63] they may disappear at low temperatures, leaving the dislocations that may have only a minor effect on the saturation current. The recovery of the breakdown voltage observed in these experiments suggests that microcracks were probably not formed. Such cracks would be expected to act as permanent microplasma sites that would seriously impair the observed breakdown voltage. Cracks may, of

course, be formed in junctions that are heavily damaged, for example, by deep scratches with a diamond needle.

It has been reported that in some cases the breakdown voltage of Si junctions may be increased by mechanical damage. This effect was

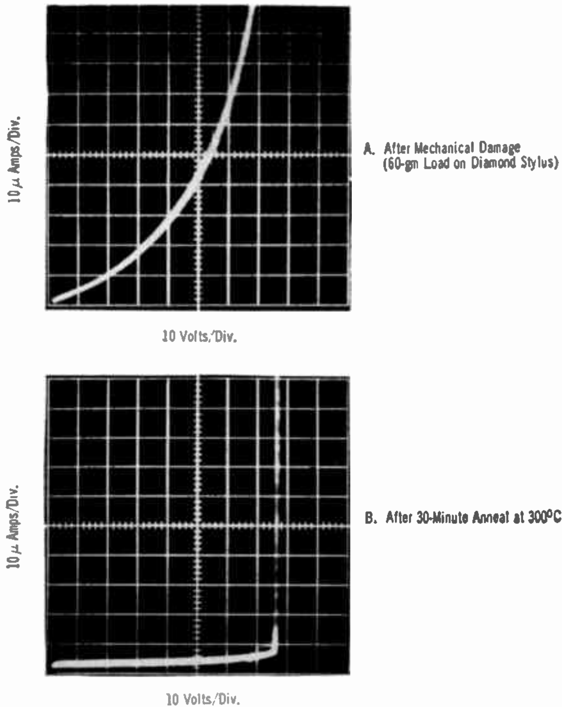


Fig. 13—Oscilloscope trace of reverse-biased silicon diode: (a) after mechanical damage (60 gm load on diamond stylus) and (b) after 30-minute anneal at 300°C (Reference (62)).

observed in n^+-p junctions and was explained by assuming that mechanical damage resulted in the introduction of a donor concentration of the order of 10^{17} cm^{-3}, thus decreasing the net carrier concentration in the p side of the junction.[64] A decrease in breakdown voltage was noted in p^+-n junctions. Substantial recovery was noted after a one hour anneal at 600°C. Detailed measurements of the change in saturation current were not given.

In addition to the permanent damage for large stresses, junctions under uniaxial stress also show a reversible change in saturation current,[65, 66] which has been explained as due primarily to a change in the band-gap energy.[62, 65]

VIII. Elimination of Microplasmas

It is evident from the preceding discussion that microplasma formation may frequently be prevented by proper precautions in the course of device fabrication, particularly by exercising the greatest care in dust prevention, cautious handling of wafers to prevent mechanical damage, and use of starting material as free of imperfections as possible. In addition, two useful methods of eliminating microplasmas are described below.

Removal of Precipitates

Because it is so difficult to prevent contamination by rapidly diffusing metals, metallic precipitates are apt to be among the chief factors in diode degradation.

A method whereby such precipitates may be removed from Si devices was described by Goetzberger and Shockley.[45] It was found that slices that would have been expected to yield soft diodes because of the presence of precipitates, partially recovered when heated at 1050°C for 30 minutes in dry nitrogen in the presence of P_2O_5 glass deposited on the surface. Further work along these lines was done by Ing *et al.*[52] These experiments were performed in planar boron-diffused p^+-n junctions, and the effect of various treatments on the reverse *I-V* characteristics of these diodes was studied. The oxides investigated (with their melting points) were B_2O_3 (577°C), P_2O_5 (563°C), V_2O_5 (690°C), Pb_2PO_7 (824°C), and the mixtures $V_2O_5 - BeO$ and $B_2O_3 - P_2O_5$. Solutions of these oxides were applied to the back of the wafers, the planar junctions having been formed in the opposite face. Heat treatment at 900°C for about one hour was found to be sufficient for the most effective oxides (P_2O_5 and B_2O_3 or a mixture of the two), yielding "sharp" breakdowns. The effect of P_2O_5 gettering was of particular interest in that the devices tended to remain stable when subsequently baked for several hundred hours at elevated temperatures (190°C) with a bias applied. The diodes gettered with the other oxides degraded after 50 hours of this treatment.

The simplest explanation for the gettering action is that the solubility of the metals in the liquid phase at the surface is larger than in the bulk of the silicon. Segregation may therefore be expected to occur. It was also suggested[45] that the metallic impurities form a chemical complex with the oxides.

In the case of diodes where the minority-carrier lifetime is deliberately reduced by the in-diffusion of impurities like gold, the prevention of precipitates is obviously difficult. As suggested earlier, one

technique consists of preventing precipitation by very rapid quenching from the diffusion temperature.

Guard-Ring Planar Structures

A guard-ring structure of the type shown in Figure 14 has been described by Batdorf *et al*[10] and Goetzberger *et al.*[11] The idea consists of surrounding the active area of the planar device (in this case a p+ -n

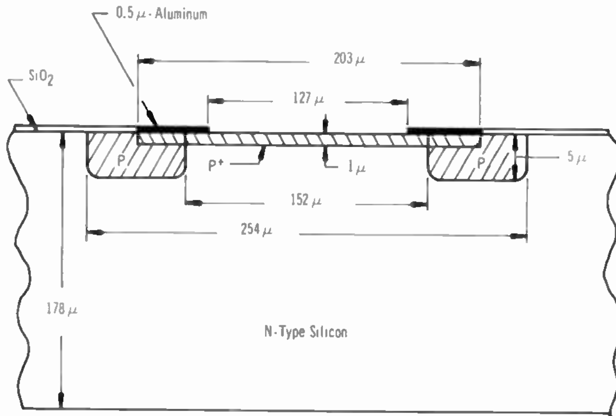


Fig. 14—Cross section of silicon guard-ring diode.

structure) with another more deeply diffused region. Breakdown should, therefore, first occur in the middle part of the device rather than at the periphery. This structure offers a convenient tool for studying diode properties free of edge effects.

Any other method that will decrease the radius of curvature at the junction edge will evidently be useful. This includes deeply diffused planar junctions. It is also of interest to note that tapering of the junction periphery has been found to be an effective method of minimizing the field at the junction periphery surface of mesa structures and hence reducing the possibility of surface breakdown.[67]

Figure 15 is a typical plot of photocurrent as a function of applied voltage of a guard-ring diode of the type shown in Figure 14. The light source consisted of a GaAs electroluminescent diode pulsed at 50 kHz. Multiplication values in excess of 300 have been measured in such devices (having active area diameters up to 500 μ) suggesting a relatively high degree of junction uniformity.[68] In similarly fabricated junctions without guard-rings, the maximum value of photocurrent multiplication that may be measured under similar conditions is generally 5 or less.

IX. Comparison of Theoretical and Observed Values of Breakdown Voltage

The ionization rates for hot carriers have been determined for Ge,[69] Si,[70, 71] GaAs,[4, 72] and GaP[73] on the basis of photomultiplication measurements. From such data it is possible to predict the avalanche

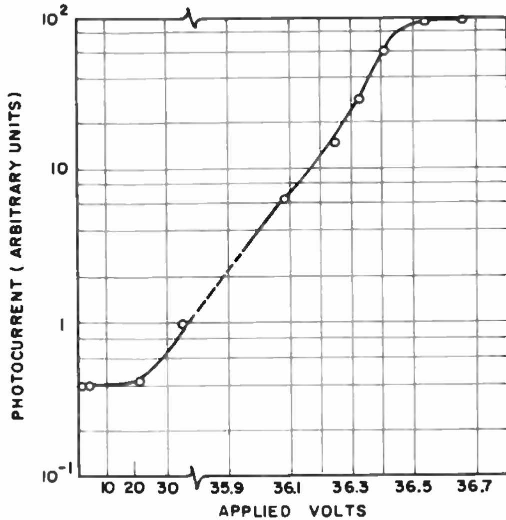


Fig. 15—Photocurrent as a function of reverse bias in a silicon guard-ring diode.

breakdown voltage of ideal p-n junctions as a function of junction parameters. It is of interest to compare such predicted values[74] with those observed on numerous samples of Si and GaAs diodes. Figure 16 compares the theoretical and observed[75] breakdown voltage for graded Si diodes as a function of impurity gradient. Figure 17 is a similar plot for abrupt Si diodes.[76] Figures 18 and 19 are similar plots for GaAs diodes. The experimental data for the GaAs junctions include both abrupt alloyed[77] and near-abrupt zinc-diffused p^+-n diodes.[78]

It is clear from these data that the typically observed values of breakdown voltage are reasonably close to those predicted on the basis of the ionization rate. A large discrepancy appears to exist, however, between the experimental and theoretical values for graded Si diodes. Since the experimental data are relatively old, it is possible that the values of breakdown voltage of these devices were strongly influenced by microplasmas. In general, good agreement between theory and experiment is improbable in the case of diodes fabricated from lightly

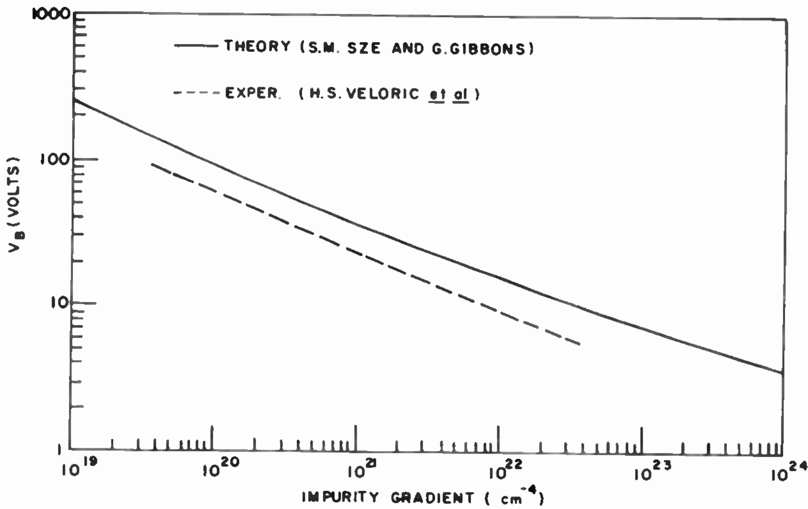


Fig. 16—Theoretical (Reference (74)) and experimental (Reference (75)) values of breakdown voltage of graded silicon diodes as a function of impurity gradient.

doped materials unless special measures are taken, such as shaped peripheries, mentioned earlier, to minimize surface breakdown.

The rather good agreement between theory and experiment over an appreciable range of breakdown voltage suggests that, despite the

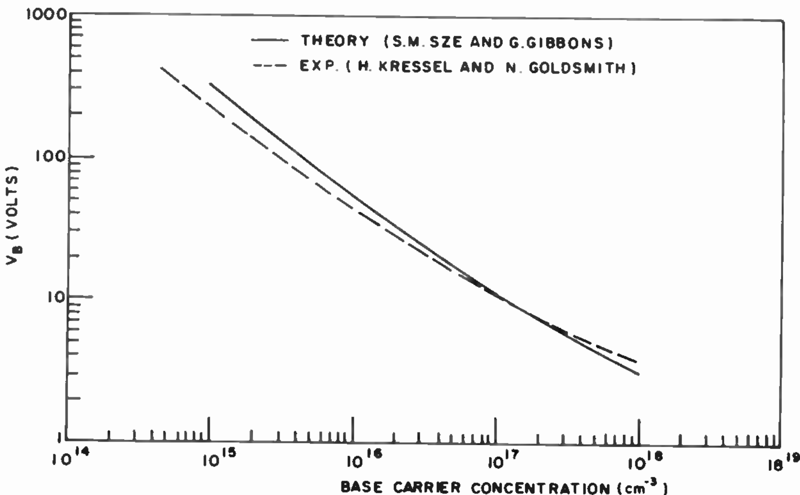


Fig. 17—Theoretical (Reference (74)) and experimental (Reference (76)) values of breakdown voltage of abrupt silicon diodes as a function of the carrier concentration of the more lightly doped base region.

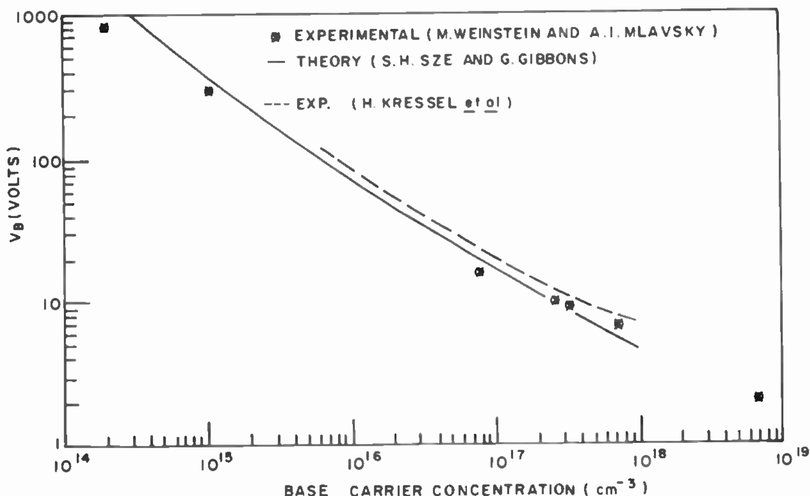


Fig. 18—Theoretical (Reference (74)) and experimental values of breakdown voltage of abrupt (References (77)) and near-abrupt zinc-diffused (Reference (78)) GaAs diodes as a function of the carrier concentration of the more lightly doped base region.

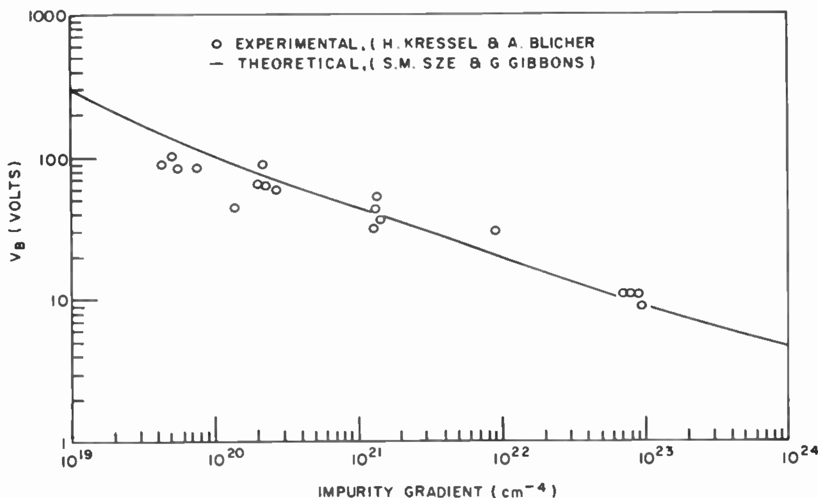


Fig. 19—Theoretical (Reference (74)) and experimental (Reference (5)) values of breakdown voltage of graded GaAs diodes as a function of impurity gradient.

presence of structural and lattice imperfections in most routinely fabricated diodes, the microplasma breakdown voltage is in general only within a few volts of the bulk breakdown, at least for the mesa devices. As discussed previously, the discrepancy for planar diodes is apt to be considerably greater, particularly for shallow junctions diffused in lightly doped substrate materials.

X. Conclusions

While a great deal of empirical information has been accumulated concerning the role of lattice defects on breakdown, the theoretical models are not well developed. In particular, the role of internal field emission at some microplasma sites, particularly at metallic precipitates, remains to be clarified. At this time, there is no theoretical model that predicts the observed current voltage behavior in junctions containing such precipitates.

So far, detailed microplasma studies have been reported only for Si devices. With the search for devices capable of operation at elevated temperatures, it becomes important to investigate the origin of microplasma sites in GaAs, GaP, and SiC junctions. The technology of GaAs is the most advanced. While it is possible to fabricate small-area (50 to 100 μ diameter) junctions having breakdown voltages reasonably close to the theoretical values (as discussed in Section IX), the fabrication of large-area rectifiers is considerably more difficult because a very much higher degree of crystalline perfection is required. In addition, it would be desirable to fabricate high-temperature rectifiers with breakdown voltages of a few hundred volts. This requires not only the ability to prepare high-purity materials but also the ability to prevent or eliminate microplasma sites.

Acknowledgments

The author is grateful to A. Elsea, G. Neighbor, and J. Preisig for technical assistance in the course of some of the experiments reported here. J. Rivera kindly made available some unpublished data. Numerous discussions with A. Blicher proved most helpful. Also gratefully acknowledged is an illuminating discussion concerning microplasmas with R. J. McIntyre and H. R. Lewis.

References

1. K. G. McKay, "Avalanche Breakdown in Silicon," *Phys. Rev.*, Vol. 94, p. 877, May 1954.

2. A. G. Chynoweth, W. L. Feldmann, C. A. Lee, R. A. Logan, G. L. Pearson, and P. Aigrain, "Internal Field Emission at Narrow Si and Ge P-N Junctions," *Phys. Rev.*, Vol. 118, p. 425, April 1960.
3. T. Tokuyama, "Zener Breakdown in Alloyed Ge P-N Junctions," *Solid-State Elect.*, Vol. 5, p. 161, Jan. 1962.
4. R. A. Logan, A. G. Chynoweth, and B. G. Cohen, "Avalanche Breakdown in GaAs P-N Junctions," *Phys. Rev.*, Vol. 128, p. 2518, Dec. 1962.
5. H. Kressel and A. Blicher, "Breakdown Voltage of Graded GaAs P-N Junctions," *J. Appl. Phys.*, Vol. 34, p. 2495, Aug. 1963.
6. G. Keil and I. Ruge, "Microplasmas in GaAs Diodes," *J. Appl. Phys.*, Vol. 36, p. 2600, Aug. 1965.
7. T. Tokuyama, "Microplasmas in Alloyed Germanium P-N Junctions," *J. Appl. Phys.*, Vol. 1, p. 324, Dec. 1962.
8. M. Gershenzon and R. M. Mikulyak, "Electroluminescence P-N Junctions in GaP," *J. Appl. Phys.*, Vol. 32, p. 1338, July 1961.
9. M. Gershenzon and A. Ashkin, "Light Emission Associated with Growth Defects from Reverse Biased GaP P-N Junctions," *J. Appl. Phys.*, Vol. 37, p. 246, Jan. 1966.
10. R. L. Batdorf, A. G. Chynoweth, G. C. Dacey, and P. W. Foy, "Uniform Silicon P-N Junctions Broad Area Breakdown," *J. Appl. Phys.*, Vol. 31, p. 1153, July 1960.
11. K. M. Johnson, "High Speed Photodiode Signal Enhancement at Avalanche Breakdown Voltage," *IEEE Trans. on Elect. Devices*, Vol. ED-12, p. 55, Feb. 1965.
12. W. Shockley, "Problems Related to P-N Junctions in Silicon," *Solid-State Elect.*, Vol. 2, p. 35, May 1961.
13. R. H. Haitz, A. Goetzberger, R. M. Scarlett, and W. Shockley, "Avalanche Effects in Silicon P-N Junctions. I. Localized Photomultiplication Studies on Microplasmas," *J. Appl. Phys.*, Vol. 34, p. 1581, June 1963.
14. A. Goetzberger, B. McDonald, R. H. Haitz, and R. H. Scarlett, "Avalanche Effects in Silicon P-N Junctions. II. Structurally Perfect Junctions," *J. Appl. Phys.*, Vol. 34, p. 1591, June 1963.
15. M. Kikuchi, "Visible Light Emission and Microplasma Phenomena in Silicon P-N Junctions," *J. Phys. Soc. Japan*, Vol. 15, p. 1822, Oct. 1960.

16. M. Kikuchi and K. Tachikawa, "Visible Light Emission and Microplasma Phenomena in Silicon P-N Junctions," *J. Phys. Soc. Japan*, Vol. 15, p. 835, May 1960.

17. J. W. Gaylord, "Microplasma Observations in Silicon Junctions Using a Scanning Electron Beam," *J. Electrochem. Soc.*, Vol. 113, p. 753, July 1966.

18. W. Czaja and J. R. Patel, "Observations of Individual Dislocations and Oxygen Precipitates in Silicon with Scanning Electron Beam Method," *J. Appl. Phys.*, Vol. 36, p. 1476, April 1965.

19. T. E. Everhart, O. C. Wells, and R. K. Matta, "Evaluation of Passivated Integrated Circuits Using the Scanning Electron Microscope," *J. Electrochem. Soc.*, Vol. 111, p. 929, Aug. 1964.

20. H. K. Henisch, *Electroluminescence*, Pergamon Press, New York, 1962.

21. A review is given by M. Gershenzon, *Physics of III-V Compounds*, p. 323, Academic Press, New York, 1966.

22. C. R. Constantinescu, G. Popovich, and P. Mihailovici, "Radiative Recombination in GaAs P-N Junctions," Conference on Luminescence, Budapest, 1966.

23. J. Shewchun and L. Y. Wei, "Mechanism for Reverse-Biased Breakdown Radiation in P-N Junctions," *Solid-State Elect.*, Vol. 8, p. 485, Aug. 1965.

24. R. Newman, "Visible Light from Si P-N Junctions," *Phys. Rev.*, Vol. 100, p. 700, Oct. 1955.

25. A. G. Chynoweth and G. L. Pearson, "Effect of Dislocations on Breakdown in Silicon P-N Junctions," *J. Appl. Phys.*, Vol. 29, p. 1103, July 1958.

26. A. G. Chynoweth and K. G. McKay, "Internal Field Emission in Silicon P-N Junctions," *Phys. Rev.*, Vol. 106, p. 418, May 1957.

27. T. Figielski and A. Torun, *Int. Conf. Phys. Semiconductors, Exeter*, p. 863, 1962.

28. K. Maeda and K. Suzuki, "Microplasmas in Silicon P-N Junctions," *Jap. J. Appl. Phys.*, Vol. 1, p. 193, Oct. 1962.

29. K. S. Champlin, "Microplasma Fluctuations in Silicon," *J. Appl. Phys.*, Vol. 30, p. 1039, July 1959.

30. R. E. Burgess, "Statistical Theory of Avalanche Breakdown in Silicon," *Can. J. Phys.*, Vol. 37, p. 730, Oct. 1959.

31. J. L. Moll, A. H. Uhlir, and B. Senitzky, "Microwave Transients from Avalanching Silicon Diodes," *Proc. IRE*, Vol. 46, p. 306, June 1958.

32. R. H. Haitz, "Studies on Optical Coupling between Silicon P-N Junctions," *Solid-State Elect.*, Vol. 8, p. 417, April 1965.

33. I. Ruge and G. Keil, "Mutual Interaction between Microplasmas in Silicon P-N Junctions," *J. Appl. Phys.*, Vol. 34, p. 3306, Nov. 1963.

34. R. J. McIntyre, "Theory of Microplasma Instability in Silicon," *J. Appl. Phys.*, Vol. 32, p. 983, June 1961.

35. R. H. Haitz, "Mechanisms Contributing to the Noise Pulse Rate of Avalanche Diodes," *J. Appl. Phys.*, Vol. 36, p. 3123, Oct. 1965.

36. H. J. Queisser and A. Goetzberger, "Microplasma Breakdown at Stair-Rod Dislocations in Silicon," *Phil. Mag.*, Vol. 8, p. 1063, Oct. 1963.

37. J. E. Lawrence and R. N. Tucker, "Tri-Pyramid and Raised-Triangle (111) Diamond-Lattice Imperfections in Silicon Epitaxial Films," *J. Appl. Phys.*, Vol. 36, p. 3095, Oct. 1965.

38. A. Goetzberger, "Investigation of Crystal Imperfections by Means of Avalanche Breakdown Patterns of Very Thin Diffused Junctions in Silicon," *Proc. Int. Conf. Semiconductor Physics, Prague, 1960*, (Academic Press, 1961).

39. S. A. Prussin in *Properties of Elemental and Compound Semiconductors*, Vol. 15 of Metallurgical Society Conferences, p. 132, (Interscience, New York, 1961).

40. J. Rivera, private communication.

41. W. Shockley and A. Goetzberger, *Properties of Elemental and Compound Semiconductors*, Vol. 15 of Metallurgical Society Conferences, p. 121, (Interscience, New York, 1961).

42. S. Prussin, "Generation and Distribution of Dislocations by Solute Diffusion," *J. Appl. Phys.*, Vol. 32, p. 1876, Oct. 1961.

43. A. W. Fischer and J. A. Amick, "Defect Structures on Silicon Surfaces after Thermal Oxidation," *J. Electrochem. Soc.*, Vol. 113, Oct. 1966.

44. J. M. Fairfield and G. H. Schwuttke, "The Influence of Crystallographic Defects on Device Performance," *J. Electrochem. Soc.*, Vol. 113, p. 1229, Nov. 1966.

45. A. Goetzberger and W. Shockley, "Metal Precipitates in Silicon P-N Junctions," *J. Appl. Phys.*, Vol. 31, p. 1821, Oct. 1960.

46. See, for example, R. G. Rhodes, *Imperfections and Active Centers in Semiconductors*, Pergamon Press, New York, 1964.

47. W. C. Dash, "Birefringence in Silicon," *Phys. Rev.*, Vol. 98, p. 1536, June 1955.
48. G. H. Schwuttke, "Study of Copper Precipitation Behavior in Silicon Single Crystals," *Electrochem. Soc.*, Vol. 108, p. 163, Feb. 1961.
49. J. E. Lawrence, "Electrical Properties of Copper Segregates in Silicon P-N Junctions," *J. Electrochem. Soc.*, Vol. 112, p. 796, Aug. 1965.
50. A. N. Knopp and R. Stickler, "Transmission Electron Microscope Investigations on Diffused Silicon Wafers with Relation to Electrical Properties of Controlled Rectifiers," *J. Electrochem. Soc.*, Vol. 111, p. 1372, Dec. 1964.
51. R. N. Hall and J. H. Racette, "Diffusion and Solubility of Copper in Extrinsic and Intrinsic Ge, Si, and GaAs," *J. Appl. Phys.*, Vol. 35, p. 379, Feb. 1964.
52. S. W. Ing, Jr., R. E. Morrison, L. L. Alt, and R. W. Aldrich, "Gettering of Metallic Impurities from Planar Silicon Diodes," *J. Electrochem. Soc.*, Vol. 110, p. 533, 1963.
53. W. Kaiser, "Electrical and Optical Properties of Heat-Treated Silicon," *Phys. Rev.*, Vol. 105, p. 1751, March 1957.
54. K. Sato, "On Si-C Contamination of Silicon Epitaxial Wafer," *Solid-State Elect.*, Vol. 7, p. 743, Aug. 1964.
55. M. Yoshida and K. Furusho, "Wirkung der Mengendichte der Ausscheidung auf den Keim der Nickelausscheidung entlang der Versetzung in Silicium," *Jap. J. Appl. Phys.*, Vol. 4, p. 74, Jan. 1965.
56. G. J. Brucker, W. J. Dennehy, and A. G. Holmes-Siedle, "Ionization and Displacement Damage in Silicon Transistors," *IEEE Trans. on Nuclear Science*, in press.
57. J. E. Drennan and D. J. Hamman, "Space Radiation Damage to Electronic Components and Materials," Radiation Effects Information Center, Battelle Memorial Institute, Columbus, Ohio, 1966.
58. J. Hilibrand, private communication.
59. C. D. Root, D. P. Lieb, and B. Jackson, "Avalanche Breakdown Voltages of Diffused Silicon and Germanium Diodes," *IRE Trans. on Elect. Devices*, Vol. ED-7, p. 257, Oct. 1960.
60. S. M. Sze and G. Gibbons, "Effect of Junction Curvature on Breakdown Voltage in Semiconductors," *Solid-State Elect.*, Vol. 9, p. 831, Sept. 1966.

61. J. Shields, "Avalanche Breakdown Voltage in Hemispherical P-N Junctions," *J. Elect. and Control*, Vol. 4, p. 58, Jan. 1958.

62. H. Kressel and A. Elsea, "Effect of Generation-Recombination Centers on the Stress-Dependence of Si P-N Junction Characteristics," *Solid State Elect.*, Vol. 10, p. 213, March 1967.

63. G. D. Watkins and J. W. Corbett, *Discussions of the Faraday Society*, Vol. 31, p. 86, Jan. 1961.

64. A. Goetzberger and R. H. Finch, "Influence of Mechanical Damage on Avalanche Breakdown in Silicon P-N Junctions," *Solid-State Elect.*, Vol. 7, p. 543, June 1964.

65. J. J. Wortman, J. R. Houser, and R. M. Burger, "Effect of Mechanical Stress on P-N Junction Device Characteristics," *J. Appl. Phys.*, Vol. 35, p. 2122, July 1964.

66. For a review of experimental results and applications see W. Ridner, G. Doering, and R. Wonson, "Structural and Operational Characteristics of Piezo-Transistors and Allied Devices," *Solid-State Elect.*, Vol. 8, p. 227, March 1965.

67. R. L. Davies and F. E. Gentry, "Control of Electric Field at The Surface of P-N Junctions, "*IEEE Trans. Elect. Devices*, Vol. ED-11, p. 313, July 1964.

68. H. Kressel and A. Elsea, unpublished data.

69. B. M. Wul and A. P. Shotov, *Solid-State Physics in Electronics and Telecommunications*, Academic Press, New York, 1960.

70. S. L. Miller, "Ionization Rates for Holes and Electrons in Silicon," *Phys. Rev.*, Vol. 105, p. 1246, Feb. 1957.

71. The most recent data are those of C. A. Lee, R. A. Logan, R. L. Batdorf, J. J. Kleinack, and W. Wiegmann, "Ionization Rates of Holes and Electrons in Silicon," *Phys. Rev.*, Vol. 134, p. A781, May 1964.

72. H. Kressel and G. Kupsky, "The Effective Ionization Rate for Hot Carriers in GaAs," *Int. J. of Elect.*, Vol. 20, p. 535, Nov. 1966. (These newer values differ appreciably from those given in Ref. (4) and are in excellent agreement with the values used in Ref. (74) to calculate theoretical breakdown values.)

73. R. A. Logan and H. G. White, "Charge Multiplication in GaP P-N Junctions," *J. Appl. Phys.*, Vol. 36, p. 3945, Dec. 1965.

74. S. M. Sze and G. Gibbons, "Avalanche Breakdown Voltages of Abrupt and Linearly Graded P-N Junctions in Ge, Si, GaAs and GaP," *Appl. Phys. Letters*, Vol. 8, p. 111, March 1966.

75. H. S. Veloric, M. B. Prince, and M. J. Eder, "Avalanche Breakdown Voltage in Silicon Diffused P-N Junctions as a Function of Impurity Gradient," *J. Appl. Phys.*, Vol. 27, p. 895, Aug. 1956.

76. H. Kressel and N. Goldsmith, "High Voltage Epitaxial GaAs Microwave Diodes," *RCA Review*, Vol. 24, p. 182, June 1963.

77. M. Weinstein and A. J. Mlavsky, "The Breakdown Voltage of GaAs Abrupt Junctions," *Appl. Phys. Letters*, Vol. 2, p. 97, March 1963.

78. H. Kressel, A. Blicher, and L. H. Gibbons, Jr., "Breakdown Voltage of GaAs Diodes Having Nearly Abrupt Junctions," *Proc. IRE.*, Vol. 50, p. 2493, Dec. 1962.

A SURVEY OF RADIATION EFFECTS IN METAL–INSULATOR–SEMICONDUCTOR DEVICES*

BY

K. H. ZAININGER‡ AND A. G. HOLMES-SIEDLE†

Summary—*Unlike the solar cell and n-p-n transistor, the MIS device does not sustain a degradation as the principal effect of exposure to nuclear radiation. Instead, the MIS device undergoes a change of operating region. The change is in the nature of a parallel shift of the characteristic curve of the device, and is produced by the trapping of radiation-excited holes within the 2000-angstrom insulator and the consequent buildup of a fixed "bulk space-charge" in the insulator. Less significant changes under radiation are variations in the shape of the characteristic curve and increased leakage current, which are effectively degradations. The authors discuss the physics of radiation effects in MIS devices and recommend methods of minimizing these effects.*

I. INTRODUCTION

UNTIL THE ADVENT of the metal–insulator–semiconductor (MIS) device, the insulator film used on the surface of silicon transistors and other devices (e.g., integrated circuits and diodes) was of interest only as an inert, passivating layer. In MIS devices, however, such as the metal–oxide–semiconductor (MOS) transistor, the film is an active, integral part of the device. The ability to grow thin, uniform and continuous insulator films makes it possible to apply a high electric field to the semiconductor surface via a metal gate evaporated on top of the insulator and, thus, to modulate the semiconductor surface conductivity. Device operation is very sensitive to the properties of the gate insulator, and considerable sophistication in the production process is required to make good MIS insulator films. The migration of minute traces of impurities through the insulator, especially at elevated temperatures and under high-field conditions, can cause serious drifting of the type shown in Figure 1. It now further appears that radiation can cause similar drift and/or degradation in device characteristics, although the drift is caused by distinctly different processes.

Investigation of these distinctions and the basic nature of the

* Presented at Symposium on Radiation Effects in Semiconductor Components, Toulouse, France, March 1967.
† RCA Astro-Electronics Division, Princeton, N. J.
‡ RCA Laboratories, Princeton, N. J.

effects of radiation on MIS devices involves the techniques and concepts of semiconductor surface physics. This review attempts to relate these concepts to their manifestations in present-day MIS devices. The importance of these effects arises from the interest in the MOS transistor as a very small, inexpensive, light, and low-power electronic device in industrial, space, and military electronic systems. In some highly specialized applications, MIS devices may be subjected to radiation levels in the megarad range. Typical radiation sources are radioisotope power

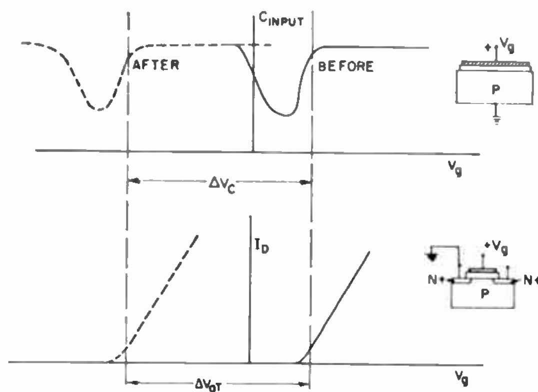


Fig. 1—Change in operating region of MIS devices produced by ionizing radiation.

generators, nuclear weapons, and the high-energy particles found in space. At radiation doses greater than a kilorad, exposure may cause strong electrical changes in the MIS system. In this review the basic physics of the radiation effects are described, the important irradiation experiments that have been performed on MIS devices are surveyed and compared, and finally some approaches for dealing with the effects are suggested.

The study describes work performed on both MIS capacitors and MIS transistors. The former, because of their ease of fabrication, simplicity of structure, and the sensitivity of their capacitance-versus-bias (C-V) and conductance-versus-bias (G-V) characteristics, constitute excellent test vehicles for basic studies that can lead to a physical model for device degradation under radiation. Transistor devices, especially mass-produced forms, are inevitably more complex; they are, however, the devices most commonly posing a problem for the engineer who has to design an electronic system that will work under radiation. Thus, while radiation effects in such complex devices are less amenable to precise physical definition, it is possible to use the knowledge derived

from capacitor experiments to analyze radiation effects in transistors and to develop ways of improving the basic device "hardness" to radiation.

II. Physics of the MIS System

General Comments on Solid–Solid Interfaces

An MIS structure consists of a semiconductor substrate covered by an insulator layer (between 50 Å and 5000 Å thick) upon which a metal electrode (gate) is deposited. Early work on semiconductor surfaces was mainly concerned with the semiconductor–vacuum and semiconductor–gas systems.[1] In the latter, fortuitous room-temperature air oxidation normally produced a 30-50 Å thick oxide film that was considered part of the *surface region*. This could thus be regarded as a narrow region of discontinuity between the two media. The physics of such surfaces was reasonably well understood and the localized electronic states associated with the surface region were called *surface states*. A distinction was made between *fast* and *slow* surface states, i.e., between states that could exchange charge with the semiconductor space-charge region rapidly or slowly, respectively. The slow states were generally thought to be at the outside of the oxide film, and the fast ones right at the boundary between the semiconductor and the oxide layer.

MIS physics is a natural extension of semiconductor surface physics into a systems of two solid–solid interfaces separated by an insulating film whose thickness is no longer negligible. Because of the presence of this film and its two surface space-charge regions, MIS physics is more complicated than semiconductor surface physics.[2] Depending on the ratio of the insulator thickness to the extent of the space-charge regions, three different situations may arise.

(1) If this ratio is very large, then there is no interaction between the metal–insulator and insulator–semiconductor interface regions, and the voltage drops linearly across the oxide except in the two (relatively small) surface space-charge regions. This is the condition usually assumed but rarely justified in device calculations.

(2) If the ratio is in the order of unity, then the two surface regions in the oxide begin to interact. In this case the two interfaces cannot be separated, and the voltage profile is a complicated function of all the parameters involved.

(3) For a ratio much smaller than unity, the integrated space charge in the oxide can be neglected with respect to the surface charge in the metal, resulting in a linear voltage variation through the oxide

film. The silicon surface potential in this case is essentially determined by the interaction between the metal and the silicon, and by the density of states at the interface.

The concepts and terminology used are mostly inherited from semiconductor surface physics dealing with a semiconductor–gas system. Thus, for example, the simple distinction between fast and slow surface states is not applicable in an MIS system. However, because of the widespread use of the term "surface states" we have retained this expression, but use it in an operational manner. In this usage, the term "effective surface states" lumps together the effect of work-function difference and both the charge configurations that can be defined for the MIS system, namely, *interface states* and *oxide charge.*

Charge Relations in the MIS System

The charge neutrality of an MIS system is frequently expressed as

$$Q_{ss} + Q_{sc} + Q_M = 0\,, \tag{1}$$

where Q_{ss}, Q_{sc}, and Q_M indicate charges in the effective surface states, semiconductor space-charge region, and metal electrode, respectively. In reality these quantities represent effective charges with considerable complexity as regards their distribution and energetics of formation.[3] Charge-exchange equilibria exist between each region, and several different types of charge can be distinguished within the insulator phase itself. A clear experimental separation of the charge in the effective surface states into its components, including their polarity, is difficult. Each measurement method can only distinguish between certain kinds of charge and, thus, classifications depend to a large extent on the method of measurement employed. Since high-frequency (~ 1 MHz) C-V measurements are relatively easily and rapidly carried out,[4] and since it has been shown[5-6] that they can yield interesting and meaningful results if the proper experimental conditions are established and if care is taken in their interpretation, it is advantageous to classify the surface states in the MIS system on the basis of this measurement method. Thus, we have the following two categories.

(1) *Interface states* are stationary electronic states located right at the (Gibbsian) plane separating the semiconductor from the insulator. Interface states are analogous to the fast surface states mentioned above. Because the insulator has a wide forbidden gap, the energy levels of interface states can lie either within or outside the forbidden gap of the semiconductor. This will determine whether or

not they change their charge state when a field is applied between metal and semiconductor.

(2) *Oxide charge* is charge that is trapped in the space-charge region of the insulator film and cannot communicate with the semiconductor surface. There are two forms of this charge, namely, mobile and immobile. The immobile species is mainly held in traps that are part of the intrinsic defect structure of the insulator; the mobile charge is mostly due to ions that are capable of migrating through the insulator, especially during conditions of high field and elevated temperature.

Measurement Methods

Several measurement methods are available that give certain distinctions between the various types of charge in the MIS system.

The High-Frequency MIS Capacitance Method

The small-signal differential admittance of an ideal MIS capacitor with no surface states consists of the insulator capacitance, C_i, in series with the surface space-charge layer capacitance,[7-11] C_{sc}. C_i is independent of frequency. C_{sc} also is independent of frequency in the accumulation and depletion regimes (up to 10^{11} Hz), but in the inversion regime, C_{sc} takes on different forms depending on whether or not the minority carriers can follow the applied a-c signal and/or bias. The dependence of this ideal MIS capacitance on effective bias is shown in Figure 2 for the case of a metal — SiO_2 — Si structure. In all cases, this capacitance is uniquely determined by the semiconductor surface potential, ψ_s. For an actual MIS diode in which surface states are present but in which loss mechanisms are neglected, the above equivalent circuit is modified by adding the surface state capacitance, C_{ss}, in parallel with C_{sc}. The dependence of C_{ss} on frequency and surface potential is a function of the density of surface states and their spatial and energy distribution. Since this is, in general, not known, little information about the physical properties of the interface can be obtained from an analysis of the MIS *C-V* characteristics obtained at an arbitrary frequency. However, if the measurement frequency is sufficiently high so that surface states cannot follow, then the surface-state capacitance becomes zero and the MIS capacitance reduces to its high-frequency form, i.e., the series combination of C_i and C_{sc}. When this condition is satisfied, the MIS capacitance is unambiguously related to the semiconductor surface potential. However, there is a difference between an experimentally determined high-frequency *C-V*

characteristic and one computed for an identical structure without surface states, and this is the voltage due to the total charge in surface states. By finding the difference, ΔV, between the measured voltage for a given capacitance value, say the "flat-band" capacitance, and its "ideal" value, one can determine the total charge that is trapped in surface states as a function of surface potential. The density of

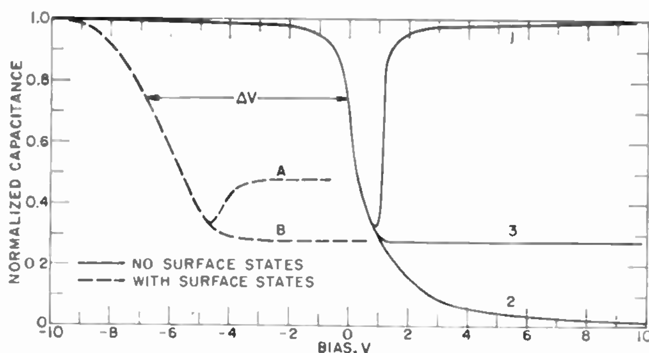


Fig. 2.—Normalized MIS capacitance vs. bias for 10 ohm-cm p-type silicon with 1000-Å SiO_2 film. The solid curve is for the case of no surface states. In the inversion regime curve (1) holds if the minority carriers follow both a-c, and d-c signal, curve (2) if they cannot accumulate at the surface (Schottky depletion layer capacitance), curve (3) if they follow d-c but not a-c signal. The dashed curve is an experimental result for an identical structure with surface states. Curve (A) corresponds to curve (1) in the ideal case except for the limited response of the minority carriers; curve (B) is analogous to curve (3). ΔV is a measure of the density of ionized surface states.

effective surface states for any particular value of capacitance, as reflected to the insulator–semiconductor interface, N_{ss}, is then given by

$$N_{ss} = C_i\,\Delta V/(1.6 \times 10^{-19}) = (\epsilon_i \Delta V)/(1.6 \times 10^{-19} d_i)\,, \qquad (2)$$

where ϵ_i and d_i are the insulator dielectric constant and thickness, respectively. If the experimental *C-V* curve is to the left of the ideal curve, the charge in surface states is positive; if it is to the right, the charge is negative. A comparison between experimental and ideal curves is shown in Figure 2.

The MIS Conductance (G-V) Method

The small-signal differential a-c conductance, G, of an MIS diode is essentially due to the exchange of charge between interface states and the bulk semiconductor. If this conductance is measured as a function of bias and frequency, information concerning the density,

distribution, time-constants, and capture cross section of the interface states present can be obtained.[12]

Channel Conductivity (I_d-V_g) Method

The property of MIS devices most commonly made use of in electronics is the modulation of channel conductivity g by gate bias V_g. Drain-to-source current I_d is, of course, controlled by g,

$$g = \frac{I_d}{V_{ds}} = \frac{W\mu C}{L}\left(V_g - V_T - \frac{V_{ds}}{2}\right), \tag{3}$$

where V_{ds} is the drain-to-source voltage, V_g is gate voltage, V_T is "pinch-off" or threshold voltage (here taken as a negative quantity as in n-channel depletion devices), μ is the carrier mobility, C is the gate-to-substrate capacitance per unit area, L is the channel length (i.e., source–drain separation), and W is the channel width. As "pinch-off" is a gradual phenomenon, the definition of V_T must include a statement as to the channel current or conductivity value chosen to represent the threshold between the turned-on and pinched-off conditions.

Under certain conditions, measurements of the shift in threshold voltage (ΔV_T) induced in an MIS transistor by radiation can be used to determine the density of charge ΔQ newly introduced into the oxide (see, e.g., Figure 8). The introduction of positive charge into oxide states and interface states induces a balancing increase of negative conduction electrons in the channel. The product of C and ΔV_T represents the amount of electronic charge that must be suppressed in the channel in order to bring the conductivity back to the defined threshold level.

Particular care is needed, however, in interpreting such data. Gate–source capacitance must be eliminated from C; it must be assumed that semiconductor doping is not changed by the irradiation; and V_{ds} must be small enough to be insignificant in Equation (3). If new interface states are generated by the irradiation (and several experiments discussed later suggest that they frequently are), the shape of the I_d-V_g characteristic will be affected, since changing the bias may now cause emptying or filling of these states. Thus, the calculated value of fixed oxide charge density from I_d-V_g characteristics will contain an uncertainty equal to the density of interface states created. Despite these limitations, the I_d-V_g characteristics is of use for studying the inversion regime, while the C-V characteristic gives more unambiguous data concerning the accumulation regime.

Other Specific Measurement Methods

The above methods give special information about electrical characteristics of insulators, but no direct structural information about the material. Other methods, such as electron spin resonance and visible-light spectroscopy give highly specific structural information. These methods are important in determining the defect structure of insulators.

III. Brief Chronology of Studies of Radiation Effects in MIS Devices

The effect of radiation on MOS devices was first noted by Hughes and Giroux.[13] This observation removed the earlier impression that, because surface-channel devices would not be affected by the production of minority-carrier recombination centers in the silicon, they would not undergo strong changes under radiation. Even though radiation-induced charge trapping had long been recognized as the source of "color center" formation[14,15] in bulk silica and other insulators, it was not appreciated that similar effects could strongly influence the performance of MIS devices.

In early 1964, unusual degradation effects were also observed in planar bipolar transistors* under irradiation. Factual reports of this effect were made by Peden[16] and Peck and Schmid[17] in 1964. Experiments that clarified the effect were made in late 1964 by Green[18] and by Brucker, Dennehy, and Holmes-Siedle.[19,20] These indicated that charge-trapping effects in the passivating oxide were almost certainly the cause of the phenomena observed. The main phenomena were strong changes in d-c current gain and collector–base reverse leakage current under ionizing radiation. It could be shown that the effect was a surface effect analagous to the effect observed in MOS transistors (1) because it could be produced by x-rays and by electrons of energy below the minimum needed for bulk silicon damage and (2) because a strong, regular dependence of the gain degradation on injection level was always observed.[20-22] During 1965, Szedon and Sandor,[23] Speth and Fang,[24] Messenger and Steele,[25] Kooi,[26] and Hughes[27] produced data on MOS devices that indicated that oxide charge build-up was clearly involved in radiation effects in grown oxides and the amount of charge trapped depended strongly on the field exerted on the oxide while it

* These were junction transistors in which the intersection of junction and surface was protected from surface contamination, and consequent leakage effects, by the oxidative growth of a "passivating layer" of about 5,000 A SiO_2.

was being irradiated. In late 1965 Zaininger published capacitor measurements that indicated that a well-defined hole-trapping effect was involved and was accompanied by variable amounts of interface state production.[28,29] Grove and Snow[30] published confirmatory capacitor data. In 1966, evidence was produced for strong interface-state production in transistors by Dennehy and coworkers,[31] and studies of the practical aspects of the use of MIS devices under radiation were made by Gordon and Wannemacher[32] and Barry and Page,[33] while further details of the effect of bias on the radiation-induced shifts in MOS and MIS transistors were reported by Stanley.[34] At present, several more profound studies of mechanisms and process effects are in progress, employing new techniques such as measurement of oxide-charging currents,[34,35] probing with fine electron beams of varied energy,[36,37] and the correlation of degree of radiation effect with the process techniques used.[38]

The conclusion the authors have drawn from their experiments are that the two most important effects occurring when MIS structures are exposed to ionizing radiation are (1) the introduction of a fixed positive oxide space charge by the occupation of pre-existing charge-trapping sites in the oxide and (2) the creation of new interface states at the silicon–silicon dioxide interface.

Recent tests indicate that atomic displacement damage in the silicon or the insulator does not play a significant role in MIS transistor operation. Effect (1) appears to occur with all insulators to some extent and to be a strong feature of amorphous insulators of very low conductivity. Effect (2) is not universal in MIS systems and the magnitude of the effect appears to depend strongly upon the initial perfection of the interface. In the following sections, each effect will be analyzed in detail.

IV. Mechanisms of Radiation-Induced Oxide Space-Charge Production

Experimental Data

General

The radiation effect in MIS devices that is universally observed is the build-up of a fixed oxide charge. If no interfering effects are present, this results in an essentially parallel shift of the *C-V* and/or *I-V* characteristic. The parallel shift in characteristics that occurs in MIS devices under radiation appears to be the most significant effect as well as being the most amenable to analysis by virtue of the static nature of the charge configuration. Figure 3 shows an example of a

parallel shift of the *C-V* and *G-V* characteristic of an MOS capacitor under irradiation. The conductivity characteristics (I_d-V_g) of the inversion layer in the semiconductor would also follow exactly these parallel shifts. The degree of parallel shift observed in MIS devices varies very widely with the nature of the insulator film. Indeed, for the normal grown silicon dioxide layers, there is even a striking dependence on the conditions used for the growth process.

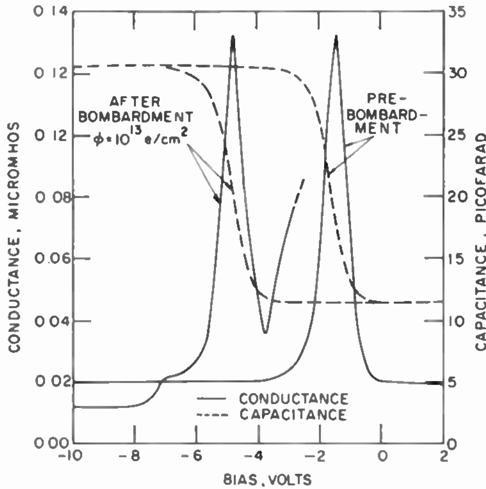


Fig. 3—*C-V* and *G-V* shifts for the same MOS capacitor before and after electron bombardment at zero gate bias.

The degree of parallel shift is also very strongly dependent on the field applied across the oxide during the particle bombardment. In this paper the applied field is usually given in terms of the bias applied to the metallic gate of the device during irradiation, and is called the bombardment bias, V_B. Figure 4 is a generalized form of the behavior of those MIS devices that undergo virtually parallel shifts. The form shown is common to most of the data published and encourages the view that a common process occurs in most MIS devices.

Several features are worth noting. There is a strong dependence of the threshold-voltage shift (ΔV_T) on bombardment bias, negative bias having a smaller effect than positive bias. At a certain dose level, saturation occurs, and this level is higher under high bias. In the worst cases observed, the arbitrary ΔV_T scale shown can be read directly in volts, while the best MIS devices show less than 1/10th this effect.

It is seen later that the gross features of this figure can be ex-

plained by some modification of an existing theory developed for hole trapping in bulk silica.

Dependence of Shift on Bombardment Bias

Figure 5 shows ΔV_T, normalized to density of charge induced,* as a function of bombardment bias for a pair of oxides grown under

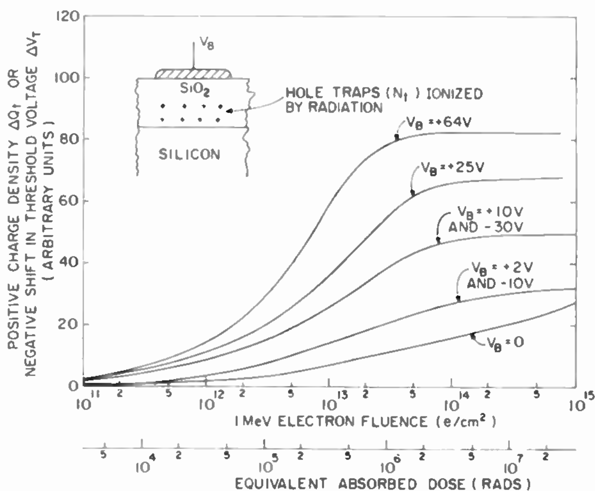


Fig. 4—General form of shift in MIS characteristics as a function of total irradiation. Gate bias during bombardment (V_B) is the parameter.

different conditions and irradiated with 1 MeV electrons to the equivalent of several megarads of x-rays. The shape of both quadrants is roughly parabolic at low bias values. Grove and Snow[30] have shown that the data fits Von Hippel's theory[39] for color-center formation in alkali halides illuminated under bias, which predicts that, for any bias value, a steady-state thickness of space charge will be achieved when all hole traps within this range (of volume density N_t) have been saturated with holes. In an ideal case,

$$d = \frac{2K_o \epsilon_o V_B^{1/2}}{qN_t}, \qquad (4)$$

* This normalization eliminates the effect of different thicknesses and dielectric constants from the comparison of insulators. However, for the two insulators shown, since they are of the same material (grown silicon dioxide) grown using different oxidant gases, and of about the same thickness (0.12μ or 1,200Å), the results can be read roughly as ΔV_T on the vertical scale with a conversion relation: ΔV_T of -10 volts $= \Delta Q_t$ of 1.76×10^{12} holes/cm^2.

where K_o = dielectric constant of the oxide, ϵ_o is the permittivity of free space, V_B is bombardment bias, and q is unit electronic charge. It can further be shown that for low bias values the saturation value for volume density of charge in the oxide (Q_{sat}), is

$$Q_{sat} = qN_t d \left(1 - \frac{d}{2x_o}\right). \tag{5}$$

where x_o is the oxide thickness.

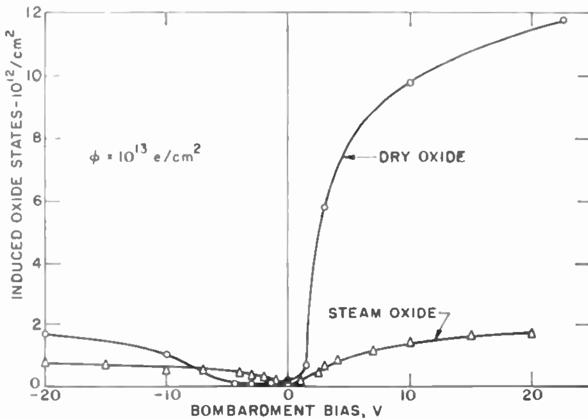


Fig. 5—Radiation-induced shifts in MOS capacitor characteristics (oxide charge generated ΔQ versus bombardment bias V_B).

Such a simple model fits the curves of Figures 4 and 5 reasonably well (see also Figure 13). Von Hippel's theory indicates that the amount of positive charge (and hence ΔV_T for a given oxide thickness) induced by a large (i.e., saturating) amount of radiation will be $Q_{sat} \propto V_B^{1/2}$. Thus, the roughly parabolic shape of the bias dependence in Figure 5 would be expected. However, other experiments demonstrate a bias dependence of ΔV_T that more nearly approaches linearity, i.e. $Q \propto V_B$. Thus, the data presently available is conflicting and probably not sufficient to generalize on the exact form of the dependence. The discrepancy probably lies in the fact that, in practice, very large amounts of radiation (in excess of 10^9 rads) are required to produce saturation of the radiation effect under all bias conditions, although, as indicated in Figure 4, the rate of change of ΔV_T with radiation dose invariably decreases very markedly after about 10^6 rads.

In addition, the assumption of hole traps at a single energy level and uniformly distributed throughout the insulator is somewhat over-

simplified and can only serve as a first approximation in gaining understanding of the radiation process. Actually, there is ample evidence that insulators have a more or less continuous distribution of defect states throughout the forbidden gap,[40] even though the density of certain levels may be large as compared with that of the quasi-continuum of states.

Furthermore, it is reasonable to assume that, depending on the method of formation of the insulator film, gradients exist in the spatial

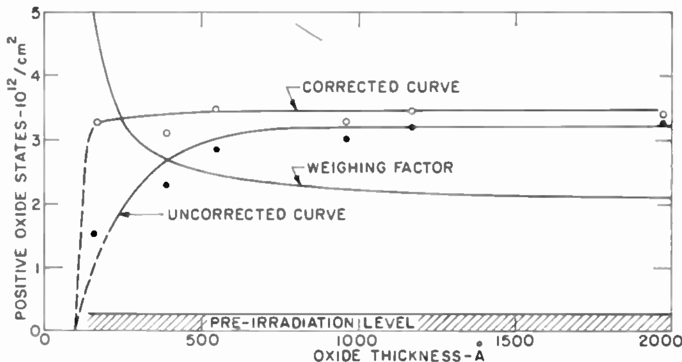


Fig. 6—Experiment on spatial location of oxide charge within oxide in dry-oxygen-grown MOS capacitor. Successive layers were dissolved by etching.

distribution of the traps. A thorough treatment of the bias dependence of the radiation damage has to take into account all the levels, and their spatial distribution. It is quite possible, however, that there will always be a discrepancy in different authors' experimental results because of the different defect structure of insulator films prepared in various ways.

Distribution of Oxide Charge

It is important to know the profile of the radiation-induced fixed oxide charge density over the depth of the oxide, since it can be demonstrated[10] that the image charge in the semiconductor channel due to a fixed charge in the oxide is a function of its distance from the oxide–silicon interface. Von Hippel's theory predicts a form of the space-charge in the oxide in which the density decreases with distance from the oxide–silicon interface. This feature of the distribution has been confirmed in experiments by one of the authors[29] and by Grove and Snow.[30] The profile was studied by stepwise removal of layers of oxide with a buffered etch, thickness determination, and measurement of the oxide charge by a probe technique. The results, given in Reference (29), are shown in Figure 6. No bias was applied and a fluence of 10^{15} 1 MeV electrons/cm^2 was used.

We can see from Figure 6 that almost all the positive charge is within 100 Å of the $SiO_2 - Si$ interface. This is the region from which radiation-excited electrons can escape from the oxide even in the absence of an applied bias. In the presence of applied bias, a deeper space-charge region is likely, as described later. Its distribution will also depend on the bias applied during irradiation. Figure 7 indicates three classes of possible charge configuration.

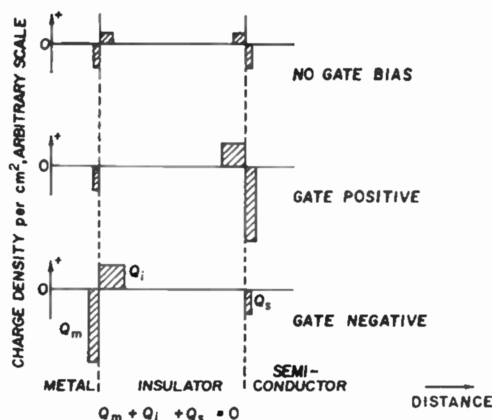


Fig. 7—A possible model for the different positive space-charge distributions produced by various values of gate bias V_B during radiation bombardment. Included also are the image charges in the metal and semiconductor.

Successive Bias-Bombardment

Several workers[29, 34] have found that the amount of charge introduced into an MIS insulator during irradiation is determined by the "fluence" (integrated flux) and bombardment bias and is almost independent of the previous bombardment history of the device. This again fits the theory of emptying or filling of existing traps. If MIS capacitors are first bombarded with various values of positive bias applied to the gate and subsequently bombarded with various negative gate biases, it is found[29] that the C-V characteristics are almost the same as those for capacitors that have been irradiated under negative bias alone. However, a small discrepancy does occur in the position of the C-V curves and raises the possibility that there may be a small amount of positive charge that is more strongly retained in the oxide.[38]

Hypothesis

The above observations have led to the physical model for fixed oxide charge production shown in Figure 8. Radiation generates elec-

tron–hole pairs in the insulator that subsequently interact with trapping sites within the insulating film. The radiation-generated electrons either recombine with the holes or move out of the insulator. The radiation-generated holes may diffuse in the insulator, but are less mobile than the electrons; many stationary hole traps are also present. The result is the observed net positive charge.

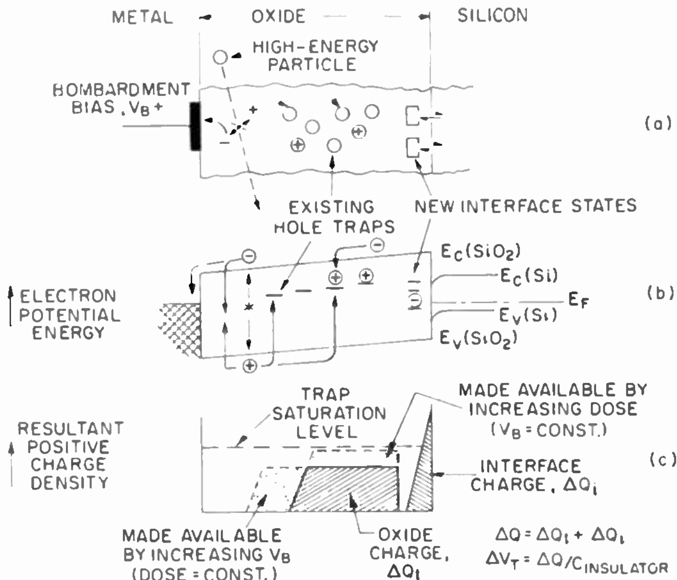


Fig. 8—Model for the effect of high-energy radiation on MIS structures: (a) device cross section, (b) energy-band picture, and (c) oxide charge distribution.

The influence of bias during bombardment on the amount and location of the radiation-induced space charge is shown in Figure 7. If there is no electric field present in the insulator during irradiation, the major effect will be electron–hole recombination; only a relatively small number of holes will be trapped and no significant charge will be introduced in the insulator. If there is an electric field present, the electrons will readily move out of the insulator film, and the trapping of holes becomes significantly enhanced. Thus, a significant positive space-charge layer is built up at the insulator–semiconductor interface and moves into the insulator with continued bombardment. Most of this charge will be imaged in the semiconductor surface and MIS measurements will show a large voltage shift.

If the gate is negative during irradiation, the positive space-charge layer is formed at the metal–insulator interface and moves into the insulator with continued irradiation. Most of the charge is imaged in the gate, and MIS measurements will not show an effect until the width of the space-charge layer is a significant portion of the insulator film. This may account for the flat minimum seen in Figure 5 at low negative biases. As Equation (4) predicts, at a given electron fluence and bombardment bias, the depth to which space charge extends into the insulator will be inversely proportional to the volume density of the charge. This depth is analogous to the depletion region in p-n junctions. This tendency is evident in Figure 5, where the influence of bias on the charge accumulated in the steam oxide begins to saturate at about 20 volts, because fewer hole traps are available and the whole film is nearly filled with space charge. On the other hand, the charge accumulated in the dry oxide is still increasing at this same bias. At high voltages, the dry-oxide curve would also presumably reach a saturation level.

If, in addition to the capture of holes, a significant trapping of electrons takes place or if the insulator is somewhat conductive (as is most probably the case for certain silicon nitride films), then no significant net space charge will be built up.

Nature of Hole Trap in Silicon Dioxide

Indications of the nature of the hole trap in amorphous oxides grown on silicon are mainly by extension from observations made on bulk silica and glass samples under irradiation. Magnetic resonance bands[41] and optical absorption bands,[42] which are common to a wide range of x-ray-irradiated glasses, have been observed. The magnetic resonances can be correlated with the dominant optical absorption. These data lead Schreurs *et al*[42] to the conclusion that: (1) the same kind of center is formed during irradiation of silicate, phosphate, and borate glasses, (2) this center is a trapped hole, and (3) the positive charge is localized on a network-forming tetrahedral group that carries a nonbridging (i.e., singly bonded) oxygen atom.

In other words, in silica derivatives, the trap is an Si-O−. Thus, the presence of a network modifier (e.g., sodium) is necessary for the production of the hole trap in glasses, since the alkali metal ions prevent the formation of a bridge between two silicon atoms. This is the reason why no optical absorption is induced by radiation in ultrapure silica. The optical absorption data indicate that, in glasses, the hole traps exist as a pair of energy levels in the forbidden gap of the insulator, lying at energies of 2.0 and 2.8 eV above the valence band.

Based on the above evidence, a tentative mechanism for the trapping of a hole in thermally grown silicon dioxide layers might be the production, by radiation, of a neutral Si-O • free radical group from an existing $Si\text{-}O^-$ group. Such species of nonbridging oxygen atom may exist even in the purest thermally grown oxides, since the formation of the oxide network in this case is a rapid process. Unlike the formation of glasses, the process is not allowed to come to equilibrium and hence some dangling bonds are likely to exist in the bulk of the oxide. With an impressed field, the ionized electron would be free to drift via the conduction band. Drift of the hole would be less easily achieved, since the availability of neighboring sites that could donate electrons to the Si-O • groups and themselves carry the charge deficit would, on the average, be very small.

It is worthy of note that, in a metal-ion-contaminated thermally grown oxide, many more $Si\text{-}O^-$ groups would exist to balance the metal ions (i.e., the metal acts as a network modifier). It would thus be expected that metal-ion-contaminated thermal-oxide films might show a larger radiation sensitivity than very pure oxide films. There is some evidence that this may be so. Other evidence that fits the extension of the bulk glass model to oxide films is the similarity of the annealing produced by heat and UV light in the two cases.

V. Production of Interface States by Irradiation

Experimental Data

In the early work on MOS capacitors,[29] it was observed that the state of the silicon oxide interface before bombardment had a strong influence on the number of interface states produced by irradiation. A result like that of Figure 3 is produced if the degree of perfection of the interface is high. If, on the other hand, the original interface already has a high density of interface states, these may be enhanced. A typical result of this type is shown in the C-V curves in Figures 9 and 10. Figure 9 shows the special case where a defect is present that gives rise to a predominance of states concentrated around a single energy level. The states at this level appear to be enhanced over and above the general background of other interface states at other energy levels. By means of further studies of effects such as this, it may be possible to assign structures to some of the important forms of interface states.

Most irradiations of mass-produced transistors involve some generation of distributed interface states, which are manifested as a change in the shape of the I_d-V_g curve and an increase in surface re-

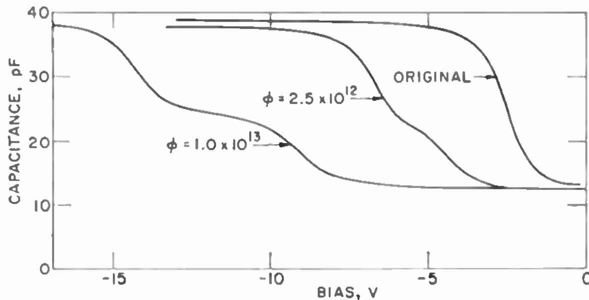


Fig. 9—*C-V* characteristics showing the effect of electron bombardment on an MIS capacitor that contains a defect that gives rise to a narrow range of energy levels for interface states.

combination velocity. In one experiment,[31] a very strong generation of interface states actually reversed the direction of shift of the transfer characteristic of an n-channel transistor. In the experiment, the room-temperature transfer characteristic also became unstable under gate bias, as shown in Figure 11. The direction of the instability is such that if positive voltage is applied to the gate, the characteristic shifts toward more positive voltage; if negative voltage is applied, the shift is toward negative voltage. These shifts are opposite in direction to the temperature-bias instabilities observed in some devices, which are attributed to the motion of ionic species in the oxide and occur only at elevated temperatures. The effect has, however, also been observed in unirradiated devices with large densities of interface states, such as many metal–silicon-nitride–silicon (MNS) devices. Kooi[26] also re-

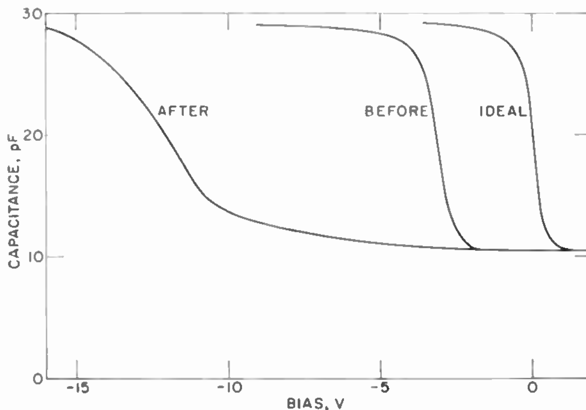


Fig. 10—*C-V* curves showing the effect of electron bombardment on an MIS capacitor that has a wide distribution of interface-state energy levels.

ported, but did not discuss, a slight instability of this kind in wet-oxygen-grown oxides irradiated by x-rays.

Hypothesis

The structure of interface states is less amenable to close definition than that of fixed oxide charge. The introduction of interface states is less regularly observed, but seems to be highly dependent on trace impurities at the surface, since impurities produce disorder[43,44] and disorder leads to a greater effect of radiation on the interface.

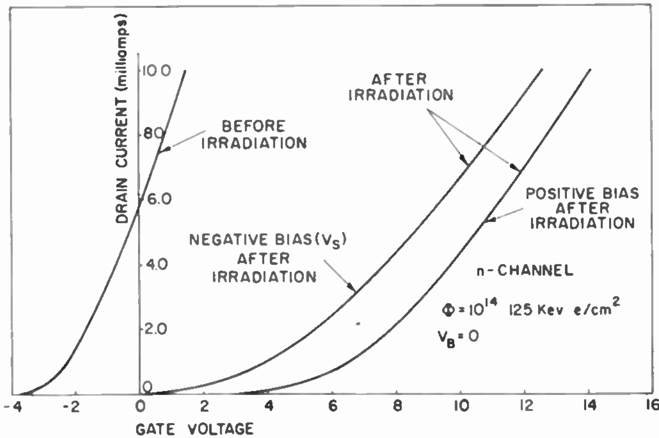


Fig. 11—Anomalous shift in characteristics of some MOS transistors under irradiation, illustrating room-temperature instability.

The radiation-induced instabilities mentioned above can be explained satisfactorily as interface states with poor communication with the interface, as described by Heiman and Warfield.[45] Further support for the hypothesis that fast surface states are created by radiation is found in measurements of surface recombination velocity at oxide-passivated surfaces. Fitzgerald and Grove found that irradiation increased the measured surface recombination velocity in planar diodes with field plates over the junction.[46,47] The authors attributed the increase to the production of fast surface states that captured colliding electrons and holes. This observation links the MIS device work with that done on radiation effects in bipolar transistors,[48,49] in which low-current gain degradation was correlated with increase in surface recombination velocity.

The fact that complete recovery of gain in bipolar transistors at 200°C is possible[20] and that interface states in MOS devices can also

be at least partially annealed[29,31,35,38] below 250°C would indicate that the formation of the states does not involve any radical change in structure at the interface.

It is not possible yet to construct a hypothesis on the structure of the radiation-induced interface states. The time constants for emptying and filling of some interface states observed can be long and, hence, some states may lie at a distance of up to 20 Å from the plane of the interface.

VI. Reversal of Radiation Effects

It is to be expected that the production of trapped oxide charge in MIS devices should be completely reversible, unlike lattice damage in silicon or the creation of interface states. This is because the only process needed for reversal is to supply the trapped hole with a recombining electron from the conduction band. This injection of electrons has been achieved by the authors and others using heat, ultraviolet light, and re-irradiation.

Temperature Annealing

Bombarded MIS capacitor samples are extremely stable at room temperature under various ambient and bias conditions, showing a reduction in oxide charge of not more than $1 \times 10^{11}/cm^2$ within a period of several months. Isochronal annealing of an irradiated MOS sample free of interface states in a dry noble gas at elevated temperatures resulted in an approximately parallel shift to the right of the *C-V* and *G-V* curves, indicating a reduction in the positive oxide charge. The major changes took place between 150°C and 300°C and were not affected by the polarity and magnitude of the gate bias during annealing. At 400°C all of the radiation effects in the oxide had been annealed out.[29,35] Similar results are observed with MOS[31] and bipolar transistors.[20]

Photo-Annealing

If SiO_2 films that have previously undergone irradiation are illuminated at room temperature by UV light, a reduction of the bombardment-induced positive oxide charge is observed.[29,35] Since UV light injects electrons from the Si valence band into the SiO_2 conduction band,[50] this experiment indicates that the positive charge induced in the oxide by the electron bombardment can be reduced or eliminated by electrons transported into the oxide from the silicon.

Radiation Annealing

A slight negative-bias bombardment of MIS devices that were initially bombarded under positive bias also largely anneals out the initial radiation effect.[29] Stanley[34] also found previous irradiation history of MOS transistors does not affect the final value of threshold voltage arrived at by irradiation of the device at a certain value of gate bias.

Discussion

Annealing at high temperature causes the trapped holes to recombine with electrons and renders the insulator essentially charge free. However, the exact mechanism for this annealing is not yet fully understood and appears to be complex, possibly involving several different processes. From the room-temperature stability of the radiation-induced positive charge one can conclude that the energy level of the hole traps must be more than about 1.5 eV above the valence band of the insulator. However, since the post-bombardment illumination of MIS capacitors with a tungsten lamp for one hour and with a UV source covered by a Pyrex filter for a half-hour[29] did not produce a shift of the *C-V* curves towards less negative voltages, i.e., did not cause a reduction of positive charge, the traps must be located more than about 4 eV above the valence band of the oxide. Snow and co-workers[35] were also unable to extract an activation energy from the observed temperature dependence of annealing.

VII. Process Effects

It has been found that insulators produced under only slightly different conditions exhibit very different behavior under radiation. In oxides, for example, growth rate, impurity content, and interface order have a major effect on sensitivity. This is understandable if the radiation sensitivity is controlled by lattice-defect structure. Thermally grown oxides, while irregular, can be considered to contain a limited number of structural configurations and a hole trap concentration in the range 10^{17} to 10^{19} traps/cm^3. Experiments concerning the relation of structure, processing and radiation effects are only of value if the oxide systems are very clean with respect to certain contaminants and have, as a result, low initial interface and oxide charges.

Oxide Growth Atmosphere

Experiments by the authors[38] have found that in very clean oxides, the presence of -OH groups appears to have a strong influence on radia-

tion sensitivity. Figure 5 illustrates the fairly general rule that oxides obtained by thermal oxidation in dry oxygen show a larger radiation sensitivity than steam-grown or anodically grown oxides, provided the degree of purity is similar. However, for each type of oxidation process, the radiation sensitivity is also influenced by post-oxidation annealing treatment. Also, it is possible to produce a range of sensitivities of dry oxides, the best of which are less sensitive than the worst steam oxides.

Surface Orientation

The dependence of the radiation sensitivity on semiconductor surface crystalline orientation has been investigated[29, 35] using capacitors made from (111) and (100) silicon with oxide grown in dry oxygen. No dependence of the radiation-induced shift on orientation could be detected. This indicates that the radiation sensitivity of MOS devices is controlled by the structure of the oxide itself, most probably its defect structure, and not by the orientation of bonds immediately at the interface.

VIII. Particle Effects

Using the model described in Section IV, in which the major effect of radiation in MOS devices is reversible electron–hole pair creation, it would be expected that the quantitative effect of different particles and particle energies would follow the ratios of their ionizing power in the oxide film. In general, this is borne out in the literature. In fact, for predictions of shifts in MOS transistors, a practical approach would be to convert particle fluxes to rads of absorbed energy (as has been done for electrons in Figure 4) and equate the effects of each on a given sample. Although some suggestions have been made that oxygen-atom motion is involved in the effects of neutrons on MIS devices,[25] no firm evidence for such an effect exists.

When the MIS system is considered in relation to its metal contacts, junction geometry and other special design features, other differences between particles may appear. One example of a particle energy dependence due to metal contacts is shown in Figure 12. These results are for an MOS capacitor made from dry-oxygen-grown oxide with various gate thicknesses. Lowering the energy of the incident beam from 1 MeV to 125 keV clearly increased the radiation effect. The increase in the effect follows the ratio of the rates of electron energy transfer to the oxide. It can be seen from Figure 12 that for 1.0-MeV electrons

the effect is largest under the 1.0-μ thick gate and smallest in the uncovered oxide, emphasizing the role played by the gate electrode.

IX. Radiation Effects on High-Performance MOS Devices Used in Electronics

General

In the course of the investigations described in earlier sections, the following MIS insulators have been irradiated; dry-oxygen grown SiO_2,

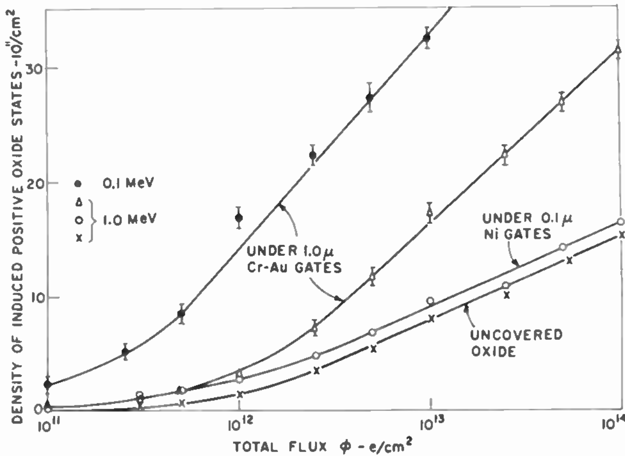


Fig. 12—Dependence of oxide charge density induced by irradiation (without bias) by electrons on 1000-angstrom oxides with various thicknesses of gate metallization.

wet-oxygen grown SiO_2, steam-grown SiO_2, anodically grown SiO_2, chemically deposited SiO_2 glasses, and silicon-nitrogen compounds ("silicon nitride"), sputtered silicon oxides and nitrides, and evaporated silicon oxides. Within each class, of course, many significant variations are possible. In the case of commercial mass-produced devices, it is not usually possible to ascertain all the important insulator fabrication parameters, and thus it may not be possible to obtain guidance on process effects from these data. However, the results of tests on some commercial devices are reviewed here in order to indicate the general level of oxide technology now in large-scale, practical use and also to point out some generically different degradation phenomena that are not observed in simpler laboratory test vehicles. Such differences may, for example, result from different channel geometry (e.g. bar geometry

versus closed geometry), special chemical treatments (e.g., phosphorus-glass coating), or greater complexity of design in contacts (e.g., there is a possibility of special effects occurring beneath long evaporated leads).

Figure 13 shows a collection of significant published experimental data on radiation-induced shifts in MOS transistors in the same form

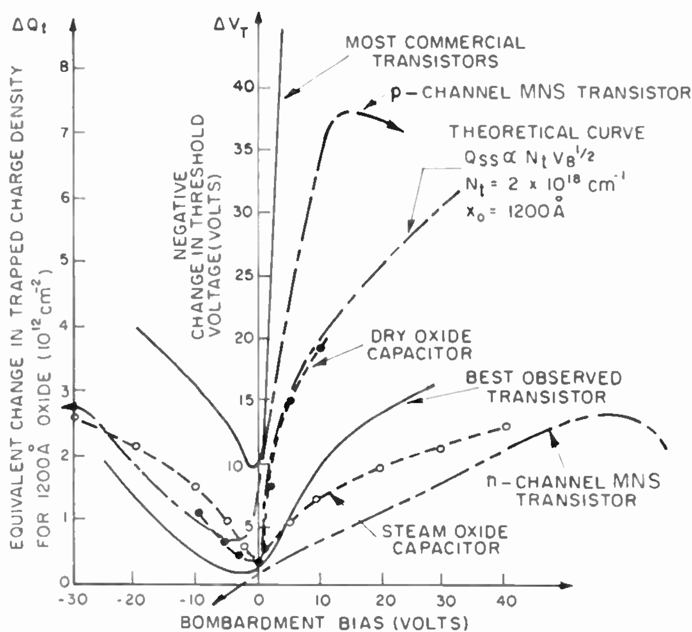


Fig. 13—Comparison of published data on radiation-induced shifts in transistor and capacitor characteristics (ΔV_T and ΔQ) versus bombardment bias (V_B).

as Figure 5. In the cases selected, the irradiation is near to that producing saturation of the radiation effect. The figure thus gives a broad indication of the limits of sensitivity found in commercial semiconductor device types in the recent past. Some curves for laboratory-grown oxides, also shown in Figure 13, indicate that lower sensitivity is definitely feasible.

The figure is plotted mainly from data on changes in I_d-V_g curves (ΔV_T). Thus, if the gate insulator thickness and dielectric constant are known, it may not be possible to relate ΔV_T to change in charge density ΔQ_t, since they are related via oxide capacitance, which includes the above terms. However, for reference, the ΔQ_t corresponding to the

given ΔV_T in an oxide 0.12 micron thick is given at the side of the figure.

In the model given in Equation (4), ΔV_T bears a square-root relation to bombardment bias, i.e., $\Delta V_T \propto V_B^{1/2}$. The theoretical curve for this relation is shown in Figure 13, and it is seen that the actual curves follow this function for some portions of their length.

The curve marked "Most Commercial Transistors" indicates the rapid rise with bias and high saturation values of the radiation-induced shifts reported by several authors for both mass-produced and laboratory-made devices.[24, 26, 34-36] This indicates that the samples quoted must have high trap concentration in the region of 3×10^{18} cm^{-2} for a 0.12-micron thickness. Thus, some feature common to most current commercial processes appears to produce high hole-trap concentrations. However, other data[29, 35, 38] on laboratory-grown MOS devices indicates that hole-trap concentrations can probably be reduced below 10^{18} cm^{-3} without resorting to very sophisticated processing techniques. Furthermore, many of the measures now being taken in the industry to ensure general device stability should also tend to reduce hole-trap concentration. A hole-trap concentration of, say, 5×10^{17} cm^{-3} would reduce the worst obtainable voltage shift for a transistor with 0.12 micron of oxide to less than 3 volts. Of course, in radiation-environment situations (e.g., within the ESSA II satellite, now orbiting in a fairly stringent environment), the effective ΔV_T would be even smaller than this because the saturation dose of radiation would not be reached for several years.

Inversion at p-n Junctions

Several workers have observed strong increases in drain-to-source leakage current (I_{DSS}) of commercial devices under bias-bombardment. Stanley,[34] for example, observed increases in the I_{DSS} value from 10^{-10} to 10^{-5} ampere for commercial devices when they were irradiated with reverse bias applied to the contact diodes. Also, after irradiation, the drain current for a strongly "on" gate bias was largely suppressed. These effects are shown in Figure 14 (inset). A generally accepted explanation for this effect, shown schematically in Figure 14, is as follows. An inversion layer develops at the edges of the contact diodes; this effectively removes the blocking diode between drain and substrate (and/or source and substrate) causing leakage. At a certain surface channel depth (represented by "a"), conduction will also be blocked along the surface channel. At higher gate biases (condition "b") conduction along the surface channel may be reinstated.

X. Operational Implications of Radiation Effects on Silicon Devices

n-Channel versus p-Channel

It is important to the electronic-circuit designer to know whether n-channel or p-channel devices are intrinsically better in any way for operation under radiation. Experience has shown that the semicon-

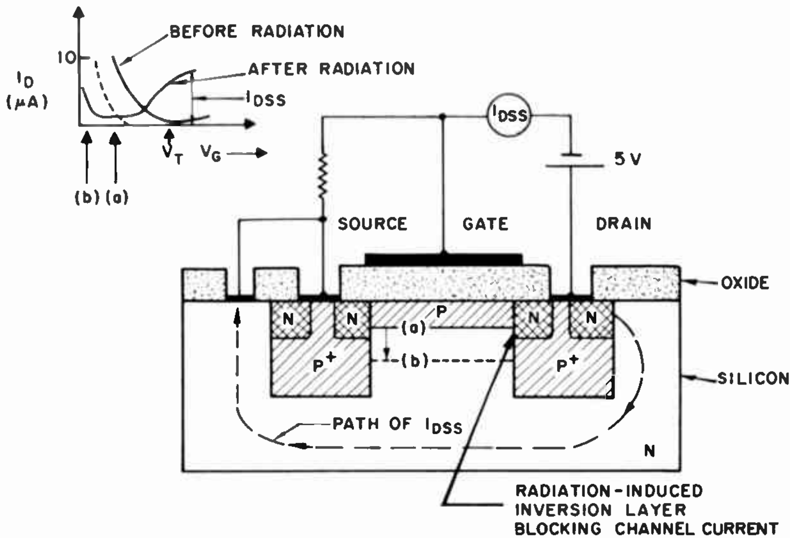


Fig. 14—Possible mechanism for increase in leakage currents from drain to substrate (or source) after irradiation. Inset shows effect of radiation sometimes seen in MOS-transistor transfer characteristics.

ductor part of an MIS system is not a strong controlling factor in determining oxide hole-trap concentration. Thus, an n-channel and p-channel transistor, made identically and irradiated under the same bias, should experience exactly the same behavior. However, the strong asymmetry of the radiation effect with respect to bombardment voltage (Figure 13) introduces an important operational factor. Logic devices that are p-channel tend to be operated with negative voltages on the gate or, at the worst, small positive voltages for the "off" portion of their duty cycle only. Thus, by virtue of the polarity of the bias, p-channel devices will exhibit smaller shifts (at the same radiation dose) than comparable n-channel devices, i.e., made by the same oxide process and operating in the analogous mode, because the latter undergoes positive excursions of gate bias. The worst case for applied bias would

seem to be the n-channel enhancement device, typically operated between two positive gate voltages (e.g., 1 and 20 volts).

MOS Logic-Circuit Effects

Commercial MOS integrated logic circuits will deteriorate in operation due to a combination of the effects described above. The dose at which such circuits fail will depend entirely on the tolerances allowed for the effects described, the monolithic circuit design itself, and the tolerance factors allowed in peripheral circuits (e.g., extra available voltage in the power supplies, small fan-out, etc). Current opinion is that, without internal changes, commercial MOS logic devices can be operated to megarad levels and higher if some drift of performance is anticipated in the circuit design and some preselection of devices for low sensitivity is carried out. Such preselection might involve (1) ensuring that units selected have not already taken up a part of the tolerance factor in performance drift by virtue of a poor fabrication step and (2) a preliminary radiation test.

Approaches to Radiation-Hardening of Oxide-Passivated Devices and Circuits

The foregoing discussions indicate that there are two general approaches to alleviating radiation effects in systems containing oxide-passivated or MIS devices: (1) changes in the device itself and (2) modifications of the system around the device in order to reduce the impact of radiation.

The second approach includes the brute-force method of shielding. This has been found practical in space because of the weak penetration properties of the radiation. In the electron and proton environment, even lightweight electronic box covers lend a significant degree of shielding. The practice will be much less rewarding in weight-limited vehicles exposed to neutron-gama sources. Other system approaches are circuit design, optimized biasing schemes, and high-temperature operation (150 to 200°C operation might produce significant thermal annealing of trapped charge).

The method of preselection has been recommended widely for bipolar transistors for use under radiation and is probably applicable to MOS devices as well. It is applicable for cases where the user has little control over the fabrication of the devices he uses. A significant amount of variation in sensitivity from batch to batch is frequently found in mass-produced, oxide-passivated devices. Indeed, significant

variations in radation sensitivity of oxides, both in MOS and bipolar devices has been observed across the surface of the same wafer of transistors. Thus, a rapid test of all devices intended for use under radiation (e.g., the actual flight units for a satellite) should be carried out, possibly using a simple radiation source such as a bench-top x-ray machine or an enclosed isotope source. Table I summarizes these approaches.

Table I—Applicability of Radiation-Hardening Techniques to MOS and Bipolar Silicon Devices

Radiation Hardening Techniques	MOS	Bipolar
Change Oxide		
(a) Reduce Impurities	x	?
(b) Modify Growth Process	x	?
(c) Use Nitride, etc.	x	?
Anneal	x	x
Preselect	?	x
Tolerant Circuit Design	x	x
Geometry Modification	x	x
Dope Bulk Silicon	—	?
Shielding	x	x

XI. Conclusions

The object of this review has been to relate the known physics of the complex MIS system to the end-use of the MIS component in military and space applications, or in other civil applications where penetrating radiation is involved.

The most universal effect of subjecting MIS structures to ionizing radiation is the generation of electron–hole pairs by interaction of the radiation with the insulator lattice, and the establishment of a positive space charge in the insulator by predominant capture of the radiation-generated holes into stationary traps. Bias during the bombardment strongly influences the number and spatial distribution of the trapped holes. Since the radiation sensitivity of MIS structures depends on the defect structure of the insulator and, thus, on the method of fabrication and annealing and/or ambient treatment, quantitative physical models can only be derived when very detailed structure and fabrication data is available and certain simplifictions are introduced into the design of test devices (e.g., low interface-state density, low impurity levels).

On the other hand, informed engineering analyses of data on less well-controlled, mass-produced circuit components may reveal important effects not evident in laboratory test vehicles. Analysis may also frequently reveal ways in which radiation-sensitive MIS devices may be used operationally in the place of other circuit elements with an overall improvement in system performance, despite the radiation effects described. In any case, it is clear from the facts presented above that the effects of radiation on MOS devices do not by any means eliminate the use of current mass-produced devices in space and other high-radiation environments so long as the circuit designer has a suitable knowledge of the relevant effects. Further, the prospects of improving MIS devices' resistance to radiation appear to be very good.

ACKNOWLEDGMENTS

The authors would like to thank Paul Rappaport for constant encouragement and interest and W. Dennehy, whose work has contributed strongly to some of the attitudes expressed in this paper.

REFERENCES

1. See, e.g., A. Many, Y. Goldstein, and N. B. Grover, *Semiconductor Surfaces*, North-Holland Publishing Co., Amsterdam, 1965.
2. See Chapters 2-4 in *Field Effect Transistors*, J. T. Wallmark and H. Johnson, Editors, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1966.
3. A. G. Revesz, "Taxonomy of Electric Charges in Metal-Insulator-Semiconductor Systems," *Proc. IEEE*, Vol. 54, p. 1002, July 1966.
4. K. H. Zaininger, "Automatic Display of MIS Capacitance Versus Bias Characteristics," *RCA Review*, Vol. 27, p. 341, Sept. 1966.
5. K. H. Zaininger and G. Warfield, "Limitations of the MOS Capacitance Method for the Determination of Semiconductor Surface Properties," *Trans. IEEE PTGED*, Vol. ED-12, p. 179, April 1965.
6. A. G. Revesz and K. H. Zaininger, "The Influence of Oxidation Rate and Heat Treatment on the Si Surface Potential in the Si-SiO_2 System," *Trans. IEEE PTGED*, Vol. ED-13, p. 246, Feb. 1966.

7. L. M. Terman, "An Investigation of Surface States at a Silicon/Silicon Oxide Interface Employing Metal-Oxide-Silicon Diodes," *Solid-State Electronics*, Vol. 5, p. 285, Sept.-Oct. 1962.

8. K. Lehovec, A. Slobodskoy, and J. L. Sprague, "Field-Effect-Capacitance Analysis of Surface States on Silicon," *Physica Status Solidi*, Vol. 3, p. 447, No. 3, 1963.

9. K. Lehovec and A. Slobodskoy, "Impedance of Semiconductor-Insulator-Metal Capacitors," *Solid-State Electronics*, Vol. 7, p. 59, Jan.-Feb. 1964.

10. A. S. Grove, B. E. Deal, E. H. Snow, and C. T. Sah, "Investigation of Thermally Oxidized Silicon Surfaces Using Metal-Oxide-Semiconductor Structures," *Solid-State Electronics*, Vol. 8, p. 145, Feb. 1965.

11. K. H. Zaininger, "Semiconductor Surface Physics," pp. 17-56 in *Field Effect Transistors*, ed. by J. T. Wallmark and H. Johnson, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1966.

12. E. H. Nicollian and A. Goetzberger, "The Si-SiO_2 Interface—The Electrical Properties as Determined by the Metal-Insulator-Silicon Conductance Technique," *Bell Syst. Tech. Jour.*, Vol. 46, 1967 (in press).

13. H. L. Hughes and R. A. Giroux, "Space Radiation Affects MOS FET's," *Electronics*, Vol. 37, p. 58, Dec. 28, 1964.

14. G. W. Arnold and W. D. Compton, "Radiation Effects in Silica at Low Temperatures," *Phys. Rev.*, Vol. 116, p. 802, Nov. 1959.

15. W. D. Compton and G. W. Arnold, "Radiation Effects in Fused Silica and $\alpha - Al_2 0_3$," *Disc. Farad. Soc.*, Vol. 31, p. 130, 1961.

16. J. Peden, IEEE Nuclear and Space Radiation Effects Conf., Seattle, Wash., July 1964.

17. S. Peck and E. Schmid, quoted in, "Surface Effects of Space Radiation on Semiconductor Devices," J. P. Mitchell and D. K. Wilson, *Bell Syst. Tech. Jour.*, Vol. 46, p. 1, Jan. 1967.

18. D. Green, J. E. Sandor, T. W. O'Keeffe, and R. K. Matta, "Reversible Changes in Transistor Characteristics Caused by Scanning Electron Microscope Examination," *Appl. Phys. Letters*, Vol. 6, p. 3, 1 Jan. 1965.

19. G. J. Brucker, W. J. Dennehy, and A. G. Holmes-Siedle, "Electron Induced Surface Damage in Silicon Transistors," *Proc. IEEE*, Vol. 53, p. 1800, Nov. 1965.

20. G. J. Brucker, W. J. Dennehy, and A. G. Holmes-Siedle, "High Energy Radiation Damage in Silicon Transistors," *IEEE Trans. PTGNS*, Vol. NS-12, p. 69, July 1965.

21. A. Holmes-Siedle, F. Liederbach, and W. Poch, "The Prediction of Space Radiation Effects on Transistors and Solar Cells," *Conference Record, 7th IEEE New York Conf. on Reliability*, May 1966 (IEEE Catalog No. F63, Paper 22).

22. M. Frank and F. Larin, "Effect of Operating Conditions and Transistor Parameters on Gain Degradation," *Trans. IEEE PTGNS*, Vol. NS-12, p. 126, Oct. 1965.

23. J. R. Szedon and J. E. Sandor, "The Effect of Low-Energy Electron Irradiation of Metal-Oxide-Semiconductor Structures," *Appl. Phys. Letters*, Vol. 6, p. 181, 1 May 1965.

24. A. J. Speth and F. F. Fang, "Effects of Low Energy Electron Irradiation on Si-Insulated Gate FET's," *Appl. Phys. Letters*, Vol. 7, p. 145, 15 Sept. 1965.

25. G. C. Messenger, E. J. Steele, and M. Neustadt, "Displacement Damage in MOS Transistors," *Trans. IEEE PTGNS*, Vol. NS-12, p. 78, Oct. 1965.

26. E. Kooi, "Effects of Ionizing Irradiations on the Properties of Oxide-Covered Silicon Surfaces," *Philips Res. Reports*, Vol. 20, p. 595, Oct. 1965.

27. H. Hughes, IEEE Conf. on Nuclear and Space Radiation Effects, Ann Arbor, Mich., July 1965.

28. K. H. Zaininger, "Electron Bombardment of MOS Capacitors," IEEE Silicon Interface Specialists Conf., Las Vegas, Nov. 1965, and *Appl. Phys. Letters*, Vol. 8, p. 140, 15 March 1966.

29. K. H. Zaininger, "Irradiation of MIS Capacitors with High-Energy Electrons," *Trans. IEEE PTGNS*, Vol. NS-13, p. 237, Dec. 1966.

30. A. S. Grove and E. H. Snow, "A Model for Radiation Damage in Metal-Oxide-Semiconductor Structures," IEEE Silicon Interface Specialists Conf., Las Vegas, Nov. 1965; *Proc. IEEE*, Vol. 54, p. 894, June 1966.

31. W. Dennehy, G. Brucker, and A. G. Holmes-Siedle, "A Radiation-Induced Instability in Silicon MOS Transistors," *Trans. IEEE PTGNS*, Vol. NS-13, p. 273, Dec. 1966.

32. F. Gordon, Jr. and H. E. Wannemacher, Jr., "The Effect of Space Radiation on MOS FET Devices and Some Application Implications of Those Effects," *Trans. IEEE PTGNS*, Vol. NS-13, p. 262, Dec. 1966.

33. A. L. Barry and D. F. Page, "Radiation Hardening of MOS Transistors for Low Ionizing Dose Levels," *Trans. IEEE PTGNS*, Vol. NS-13, p. 255, Dec. 1966.

34. A. G. Stanley, "Comparison of MOS and Metal-Nitride-Semiconductor Insulated Gate Field-Effect Transistors Under Electron Irradiation," *Trans. IEEE PTGNS*, Vol. NS-13, p. 248, Dec. 1966.

35. E. H. Snow, A. S. Grove, and D. J. Fitzgerald, Scientific Reports No. 1 and 2, USAF Contract No. AF19(628)5747, June and Dec. 1966.

36. N. C. McDonald and T. E. Everhart, Univ. Calif., Berkeley, Report No. ERL 66-16, Sept. 1966.

37. A. G. Stanley, "Effect of Electron Irradiation on Insulated Gate and Junction Gate Field Effect Transistors," IEEE Silicon Interface Specialist Conference, Las Vegas, March 1967, and Symposium on Radiation Effects on Semiconductor Components, Toulouse, March 1966.

38. W. Dennehy, K. H. Zaininger, and A. G. Holmes-Siedle, to be published.

39. A. von Hippel, E. P. Gross, J. G. Gelatis, and M. Geller, "Photocurrent Space-Charge Buildup and Field Emission in Alkali Halide Crystals," *Phys. Rev.*, Vol. 91, p. 568, Aug. 1953.

40. A. Rose, *Concepts in Photoconductivity and Allied Problems*, Interscience Publishers, London, 1963.

41. Y. Nishi, "Electron Spin Resonance in SiO_2 Grown on Silicon," *Jap. Jour. Appl. Phys.*, Vol. 5, p. 333, April 1966.

42. J. W. H. Schreurs, J. S. Stroud, and R. F. Tucker, "Charge Trapping and the Electronic Structure of Glass," VIIth Int. Congress on Glass, Brussels, Belgium, July 1965.

43. A. G. Revesz, K. H. Zaininger, and R. J. Evans, "Interface States and Interface Disorder in the Si-SiO_2 System," *Jour. Phys. Chem. Solids*, Vol. 28, p. 197, Feb. 1967.

44. B. E. Deal, E. H. Snow, and A. S. Grove, "Properties of the Silicon Dioxide-Silicon System," *SCP and Solid State Technology*, Vol. 9, p. 25, Nov. 1966.

45. F. Heiman and G. Warfield, "The Effect of Oxide Traps on the MOS Cacapitance," *Trans. IEEE PTGED*, Vol. ED-12(4), p. 167, April 1965.

46. A. S. Grove and D. J. Fitzgerald, "Surface Effects on P-N Junctions: Characteristics of Surface Space-Charge Regions Under Non-Equilibrium Conditions," *Solid-State Electronics*, Vol. 9, p. 783, 1966.

47. D. J. Fitzgerald and A. S. Grove, "Radiation Induced Increase in Surface Recombination Velocity of Thermally Oxidized Silicon Structures," *Proc. IEEE* (Letters), Vol. 54, p. 1601, Nov. 1966.

48. G. Brucker, W. Dennehy, and A. Holmes-Siedle, "Ionization and Displacement Damage in Silicon Transistors," *Trans. IEEE PTGNS*, Vol. NS-13, p. 188, Dec. 1966.

49. G. Brucker, "Correlation of Radiation Damage in Silicon Transistors Bombarded by Electrons, Protons, and Neutrons," Symposium on Radiation Effects in Semiconductor Components, Toulouse, Franch, March 1967.

50. R. Williams, "Photoemission of Electrons from Silicon into Silicon Dioxide," *Phys. Rev.*, Vol. 140, p. A569, Oct. 1965.

TIME-DOMAIN ANALYSIS OF MULTIPLE PARALLEL TRANSMISSION LINES

By

H. Amemiya

RCA Defense Electronic Products
Camden, N. J.

Summary—*With the advancement of compact packaging and rapid response of semiconductors, the study of wave propagation on multiple parallel transmission lines has become increasingly important. Published works have been limited to consideration of two-conductor transmission lines—some make the additional assumption of loose coupling. This paper presents a new method for time-domain analysis of multiple parallel transmission lines without these limitations. In the process, matched termination network and equivalent circuit concepts are developed for multiple parallel transmission lines. The new method is then applied to the case of two coupled transmission lines.*

Introduction

ELECTROMAGNETIC propagation on multiple parallel transmission lines has been a topic of continuous interest since the 1920's. An early analysis[1] was made on such lines with modified line-to-line coupling characteristics that were artificially introduced by line transpositions to obtain favorable propagation behavior. Works[2,3] published in the 1930's contained some general analyses of multiconductor transmission lines. Starting in the 1940's, a more concerted effort was directed toward the study of two parallel transmission lines.[4-9] The structure exhibited a directional coupling effect under

---

[1] J. R. Carson and R. S. Hoyt, "Propogation of Periodic Currents over a System of Parallel Wires," *Bell System Tech. Jour.*, Vol. 6, no. 3, p. 495, July 1927.

[2] L. V. Bewley, *Travelling Waves on Transmission Systems*, Chapter 6, Wiley and Sons, Inc., New York, 1933 (Second Edition 1951).

[3] L. A. Pipes, "Matrix Theory of Multiconductor Transmission Lines," *Phil. Mag.*, Vol. 24, p. 97, July 1937.

[4] M. Cotte, "Théorie de la Propagation d'Ondes de Choc sur Deux Lignes Parallèles," *Revue Générale de l'Eectricite*, Vol. 56, p. 343, Aug. 1947.

[5] M. Cotte, "Ondes de Courant sans Tension et de Tension sans Courant sur Deux Lignes Parallèles," *L'Onde Electrique*, Vol. 34, p. 381, April 1954.

[6] B. M. Oliver, "Directional Electromagnetic Couplers," *Proc. IRE*, Vol. 42, No. 11, p. 1686, Nov. 1954.

[7] W. L. Firestone, "Analysis of Transmission Line Directional Couplers," *Proc. IRE*, Vol. 42, No. 10, p. 1529, Oct. 1954.

[8] R. C. Knechtli, "Further Analysis of Transmission Line Directional Couplers," *Proc. IRE*, Vol. 43, No. 7, p. 867-869, July 1955.

[9] E. M. T. Jones and J. T. Bolljahn, "Coupled-Strip-Transmission-Line Filters and Directional Couplers," *IRE Trans. Microwave Theory and Techniques*, Vol. MTT-4, No. 2, p. 75, April 1956.

certain terminal conditions and was the topic of many papers in the microwave field.

In the present decade, with the advancement of compact packaging technique and improved rise and fall times of semiconductor devices, the study of multiple parallel transmission lines has become increasingly important. For example, the backplane wiring of a modern high-speed computer is best represented by multiple parallel transmission lines. A method is needed for time-domain analysis of such lines under general conditions. However, the works published to date have dealt with only the limited case of two conductors.[10-15] Also, some works make the additional assumption of loose coupling between lines.

The purpose of this paper is to present a new method for time-domain analysis of multiple parallel transmission lines without the limitations of previous works. Before describing the technique, the new concepts of matched termination network and equivalent circuit for multiple parallel transmission lines are described. The new approach is then applied to the case of two coupled transmission lines to obtain the same conclusions described in the previous works, although its applicability is not limited to this one case.

Matched Termination Network

A system of n parallel transmission lines and a ground generally has n modes of transverse electric and magnetic (TEM) wave propagation.[1] When only one propagation mode exists, a set of n separate matching impedances can be used, as shown in Figure 1(a), to terminate the lines without producing reflections. However, each propagation mode normally requires a different set of matching impedances for proper termination.

The matched termination network shown in Figure 1(b), on the

[10] W. J. Getsinger, "Analysis of Certain Transmission-Line Networks in the Time Domain," *IRE Trans. Microwave Theory and Techniques*, Vol. MTT-8, No. 3, p. 301, May 1960.

[11] D. B. Jarvis, "The Effects of Interconnections on High-Speed Logic Circuits", *IEEE Trans. of Electronic Computers*, Vol. EC-12, No. 5, p. 476, Oct. 1963.

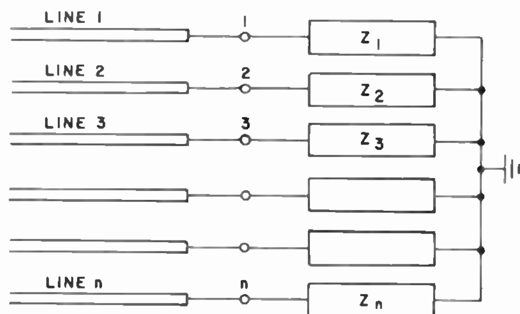
[12] H. J. Gray, *Digital Computer Engineering*, Chapter 9, Prentice Hall, Englewood Cliffs, N. J., 1963.

[13] A. Feller, H. R. Kaupp, and J. J. DiGiacomo, "Crosstalk and Reflections in High Speed Digital Systems," *AFIP Conference Proceedings*, Vol. 27, p. 511, 1965 Fall Joint Computer Conference.

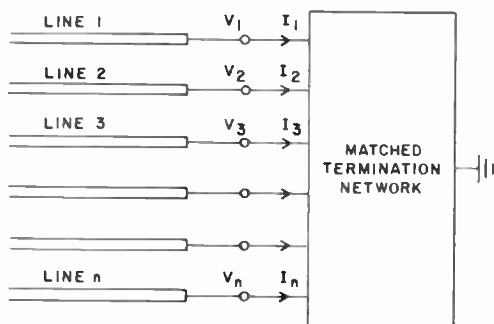
[14] N. C. Arvanitakis, J. T. Kolias, and W. Radzelovage, "Coupled Noise Prediction in Printed Circuit Boards for a High-Speed Computer System," *7th International Electronic Circuit Packaging Symposium Record (WESCON 66) IECP 2/6*, Aug. 1966.

[15] H. R. Kaupp, "Pulse Crosstalk Between Microstrip Transmission Lines," *7th International Electronic Circuit Packaging Symposium Record (WESCON 66) IECP 2/5*, Aug. 1966.

other hand, is capable of providing matched termination for any propagation mode. Since the principle of superposition holds, an arbitrary propagation involving multiple modes will be absorbed by the same matched termination network with no reflections. In other words, the



(a)



(b)

Fig. 1—Matched termination for a system of n parallel transmission lines and a ground: (a) Set of n impedances for a particular propagation mode, (b) Matched termination for any propagation mode.

matched termination network is, by definition, equivalent to the transmission lines extending to infinity.[16]

It is not always possible to build a matched termination network with a finite number of lumped constant elements. However, when the lines are lossless, a matched termination network can be constructed using $n(n+1)/2$ resistors or less in the configuration shown in Figure

[16] H. Amemiya, "Matched-Termination Network for Multiple Parallel Transmission Lines," *Electronics Letters*, Vol. 3, No. 1, p. 13, Jan. 1967.

2. One convenient way to specify a matched termination network is with the conductance matrix $[G]$:

$$[G] = \begin{bmatrix} G_{11} & G_{12} & G_{13} & \ldots & G_{1n} \\ G_{21} & G_{22} & G_{23} & \ldots & G_{2n} \\ G_{31} & G_{32} & G_{33} & \ldots & G_{3n} \\ & & \cdots\cdots & & \\ G_{n1} & G_{n2} & G_{n3} & & G_{nn} \end{bmatrix} = [I][V]^{-1},$$

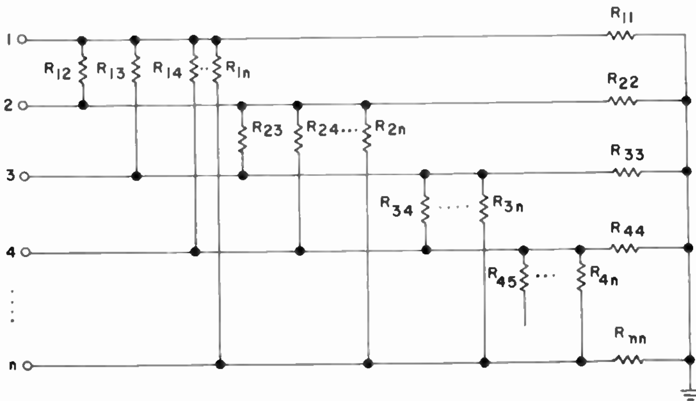


Fig. 2—Matched-termination network for lossless parallel transmission lines.

where (see Figure 1(b))

$$[I] = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ \cdot \\ \cdot \\ \cdot \\ I_n \end{bmatrix} \qquad [V] = \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ \cdot \\ \cdot \\ \cdot \\ V_n \end{bmatrix}.$$

The conductance matrix can be expressed in terms of transmission line parameters as follows (see Appendix I for proof):

$$[G] = [C]\,[V]_m\,[v]\,[V]_m^{-1} = [L]^{-1}\,[V]_m \left[\frac{1}{v}\right] [V]_m^{-1}, \quad (1)$$

where the matrixes $[L]$, $[C]$, $[V]_m$, and $[v]$ are given in Appendix I.

$[G]$ is a symmetric matrix, i.e., the network is physically realizable. Once the conductance matrix is specified, it is easy to obtain a matched termination network in the form shown in Figure 2.

$$[G] = \begin{bmatrix} \sum_{i=1}^{n} \frac{1}{R_{1i}} & -\frac{1}{R_{12}} & -\frac{1}{R_{13}} & \cdots & -\frac{1}{R_{1n}} \\ -\frac{1}{R_{21}} & \sum_{i=1}^{n} \frac{1}{R_{2i}} & -\frac{1}{R_{23}} & \cdots & -\frac{1}{R_{2n}} \\ -\frac{1}{R_{31}} & -\frac{1}{R_{32}} & \sum_{i=1}^{n} \frac{1}{R_{3i}} & \cdots & -\frac{1}{R_{3n}} \\ \vdots & \vdots & \vdots & & \vdots \\ -\frac{1}{R_{n1}} & -\frac{1}{R_{n2}} & -\frac{1}{R_{n3}} & \cdots & \sum_{i=1}^{n} \frac{1}{R_{ni}} \end{bmatrix}$$

R_{ii} = Resistance between terminal i and ground, $i = 1, 2, 3, \ldots, n$

R_{ij} = Resistance between terminal i and terminal j, $i \neq j$ and $i, j = 1, 2, 3, \ldots, n$

$R_{ij} = R_{ji}\ i, j = 1, 2, 3, \ldots, n.$

Therefore

$$R_{ii} = \frac{1}{\sum_{j=1}^{n} G_{ij}} \qquad i = 1, 2, 3, \ldots, n$$

$$R_{ij} = -\frac{1}{G_{ij}} \qquad i, j = 1, 2, 3, \ldots, n\,.$$

If the propagation velocities are identical for all propagation modes, a matched termination network can be formed by applying the following simple rule. Connect a resistor between two terminals when capacitive coupling exists between the two lines corresponding to these terminals (ground is considered to be one of the terminals). The resistance value should be the reciprocal of the product of the propagation velocity and the corresponding capacitance per unit length. (See Appendix I for proof.)

Application of the rule is shown in Figure 3. A matched termination network is obtained by substituting a resistor for each capacitance such that

$$R_{ij} = \frac{1}{v\,C_{ij}} \qquad i, j = 1, 2, 3, \ldots, n$$

v = propagation velocity

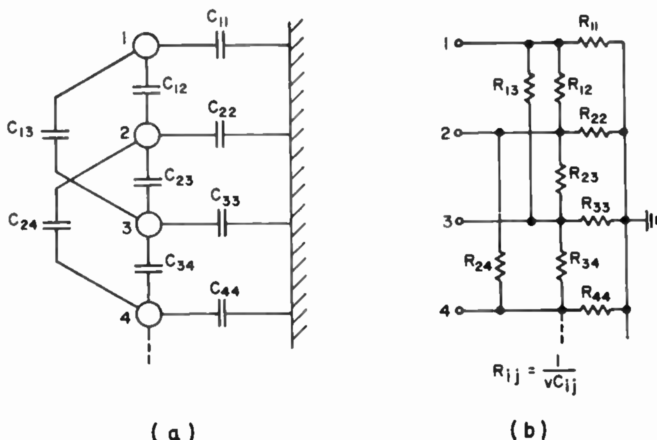


Fig. 3—Formation of a matched-termination network when propagation velocities are identical: (a) multiple parallel lossless transmission lines with mutual coupling, (b) matched-termination network.

C_{ii} = Capacitance per unit length between line i and ground, $i = 1, 2, 3, \ldots, n$

C_{ij} = Capacitance per unit length between line i and line j, $i \neq j$ and $i, j = 1, 2, 3, \ldots, n$

$C_{ij} = C_{ji} \quad i, j = 1, 2, 3, \ldots, n$.

In general, there are several equivalent configurations of a matched termination network. However, one configuration may be much simpler than others (see Figure 4). Figure 4(a) is a matched termination for n parallel lines and a ground where all lines are alike and all mutual couplings between any two lines are alike. The network is a special case of Figure 2, and consists of $n(n+1)/2$ resistors. Figure 4(b) is equivalent to Figure 4(a) in every respect, but is simpler because it consists of only $(n+1)$ resistors.

Also, if certain propagation modes are not excited, simplification of the network is possible. Using the same example, if the ground-return

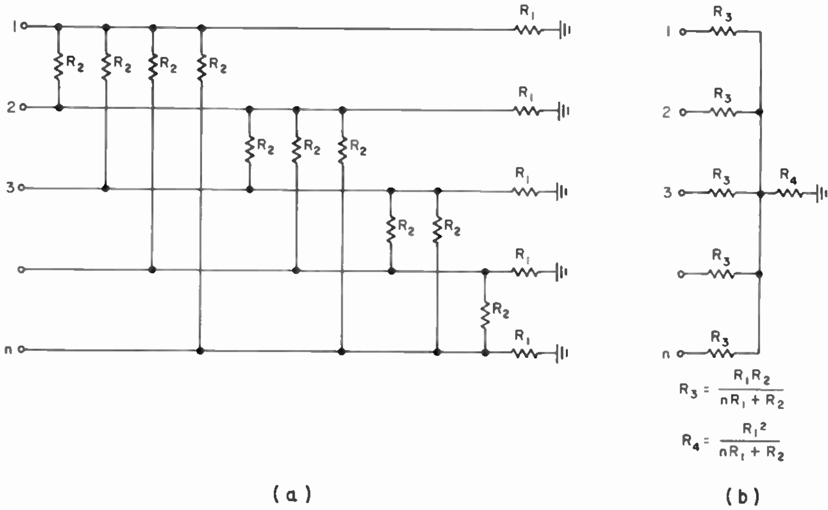


Fig. 4—Simplification of matched-termination network: (a) matched-termination network, (b) simplified network.

mode is not excited, it is possible to eliminate R_1 (n resistors) from Figure 4(a) and R_4 from Figure 4(b).

As an example, a matched termination network for two lossless parallel transmission lines with a ground will be derived (see Figure 5). The line parameters are as follows:

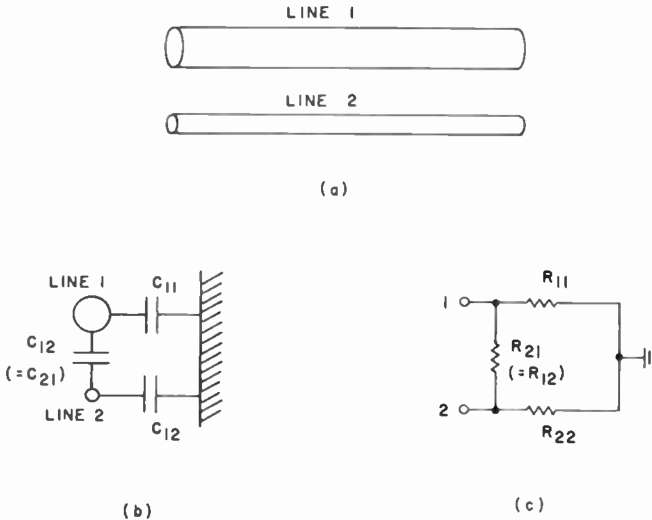


Fig. 5—Two parallel transmission lines: (a) line 1 and line 2, (b) capacitance parameters, (c) matched-termination network.

L_{11} = self inductance per unit length of line 1
L_{22} = self inductance per unit length of line 2
$L_{12} = L_{21}$ = mutual inductance per unit length between line 1 and line 2
C_{11} = capacitance per unit length between line 1 and ground
C_{22} = capacitance per unit length between line 2 and ground
$C_{12} = C_{21}$ = capacitance per unit length between line 1 and line 2

In this general case, Equation (1) is used to obtain the network. The resistance values of the termination network of Figure 5(c) are

$$R_{11} = \frac{1}{G_{11} + G_{12}}, \quad R_{12} = -\frac{1}{G_{12}}, \quad R_{22} = \frac{1}{G_{12} + G_{22}},$$

where

$$G_{11} = \frac{v_1 + v_2}{2}(C_{11} + C_{12}) + \frac{v_1 - v_2}{2D}[(C_{11} + C_{12})\{L_{11}(C_{11} + C_{12}) - L_{22}(C_{22} + C_{12})\} - 2C_{12}\{-L_{22}C_{12} + L_{12}(C_{11} + C_{12})\}],$$

$$G_{12} = \frac{v_1 + v_2}{2}(-C_{12}) + \frac{v_1 - v_2}{2D}[-C_{12}\{L_{11}(C_{11} + C_{12}) + L_{22}(C_{22} + C_{12})\} + 2L_{12}(C_{11} + C_{12})(C_{22} + C_{12})],$$

$$G_{22} = \frac{v_1 + v_2}{2}(C_{22} + C_{12}) + \frac{v_1 - v_2}{2D}[(C_{22} + C_{12})\{-L_{11}(C_{11} + C_{12}) + L_{22}(C_{22} + C_{12})\} - 2C_{12}\{-L_{11}C_{12} + L_{12}(C_{22} + C_{12})\}],$$

$$D = [\{L_{11}(C_{11} + C_{12}) - L_{22}(C_{22} + C_{12})\}^2 + 4\{-L_{11}C_{12} + L_{12}\}(C_{22} + C_{12})\}\{-L_{22}C_{12} + L_{12}(C_{11} + C_{12})\}]^{1/2}$$

$$v_1 = \sqrt{\frac{2}{L_{11}(C_{11} + C_{12}) + L_{22}(C_{22} + C_{12}) - 2L_{12}C_{12} + D}}$$

$$v_2 = \sqrt{\frac{2}{L_{11}(C_{11} + C_{12}) + L_{22}(C_{22} + C_{12}) - 2L_{12}C_{12} - D}}$$

If velocities v_1 and v_2 are equal, the following relationships must hold:

$$- L_{11} C_{12} + L_{12} (C_{22} + C_{12}) = 0$$

$$- L_{22} C_{12} + L_{12} (C_{11} + C_{12}) = 0$$

$$L_{11} (C_{11} + C_{12}) = L_{22} (C_{22} + C_{12})$$

By substituting these relationships in the preceding formulas, we obtain

$$G_{11} = v (C_{11} + C_{12}), G_{12} = - v C_{12}, G_{22} = v(C_{22} + C_{12}),$$

$$v = \sqrt{\frac{2}{L_{11} (C_{11} + C_{12}) + L_{22} (C_{22} + C_{12}) - 2 L_{12} C_{12}}}$$

or

$$R_{11} = \frac{1}{v C_{11}}, R_{12} = \frac{1}{v C_{12}}, R_{22} = \frac{1}{v C_{22}}.$$

This result could have been obtained directly by applying the simple rule given previously.

When the two lines have identical geometry, the matched termination network can be obtained by inspection. In this case,

$L_s = L_{11} = L_{22} =$ self inductance per unit length of each line

$L_m = L_{12} = L_{21} =$ mutual inductance per unit length between the two lines

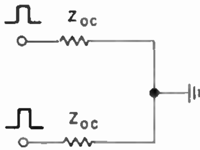
$C_s = C_{11} = C_{22} =$ capacitance per unit length between each line and ground

$C_m = C_{12} = C_{21} =$ capacitance per unit length between the two lines.

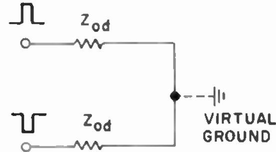
The two propagation modes are generally called the common mode (or even mode) and the differential mode (or odd mode). In common-mode propagation, the waves on the two lines are identical. In differential-mode propagation, the waves on the two lines are equal in amplitude but opposite in polarity. Each mode has its own characteristic impedance and propagation constant. For lossless lines, as assumed here, the characteristic impedances are resistive and the propagation con-

stants are imaginary, indicating that the wave propagation involves delay but not attenuation:

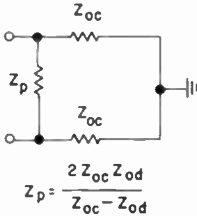
$$Z_{oc} = \sqrt{\frac{L_s + L_m}{C_s}} \text{ common-mode characteristic impedance}$$



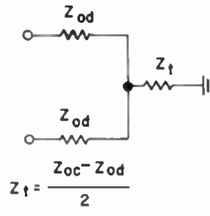
(a)



(b)



(c)



(d)

Fig. 6—Termination network for two balanced parallel lines: (a) common-mode termination, (b) differential-mode termination, (c) π matched-termination network, (d) T matched-termination network.

$$Z_{od} = \sqrt{\frac{L_s - L_m}{C_s + 2\,C_m}} \text{ differential-mode characteristic impedance}$$

$$v_1 = 1/\sqrt{(L_s + L_m)C_s} \text{ common-mode propagation velocity}$$

$$v_2 = 1/\sqrt{(L_s - L_m)\,(C_s + 2\,C_m)} \text{ differential-mode propagation velocity}$$

These characteristic impedances are defined relative to the ground as shown in Figures 6(a) and 6(b). It is a simple matter to construct networks that behave as in Figure 6(a) for the common-mode input and as in Figure 6(b) for the differential-mode input. The π network of Figure 6(c) and the T network of Figure 6(d) have such properties, and therefore are the desired matched termination networks. Normally $Z_{oc} \geqq Z_{od}$, making Z_p and Z_t positive; this condition is necessary

for physical realizability. (Figure 6(c) could have been derived from the formulas for the general case of two parallel lines.)

EQUIVALENT CIRCUITS FOR MULTIPLE PARALLEL TRANSMISSION LINES[17]

The terminal conditions of a single transmission line over a ground can be represented by an equivalent circuit consisting of a voltage or

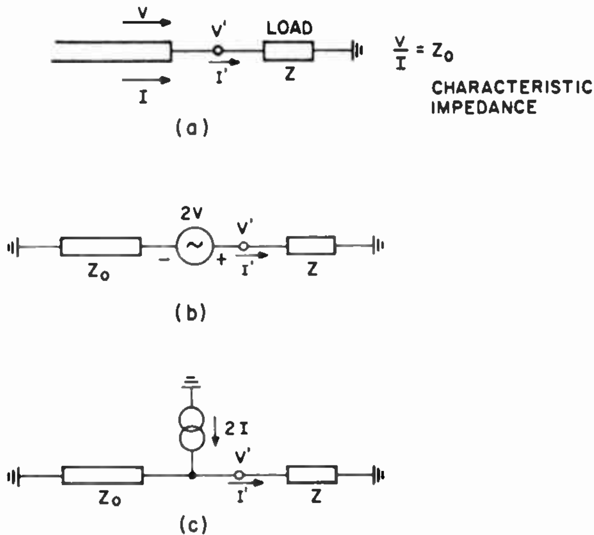


Fig. 7—Single transmission line and its equivalent circuits: (a) single transmission line terminated into a load, (b) Thevenin's equivalent circuit, (c) Norton's equivalent circuit.

current source and an impedance (Figure 7). Figure 7(a) shows a transmission line with incident voltage wave V terminated into an arbitrary load impedance Z. Voltage wave V is accompanied by current wave I and their ratio is equal to characteristic impedance Z_o. V' (terminal voltage) and I' (terminal current) are to be determined. Figure 7(b) is Thevenin's equivalent circuit with impedance Z_o and a voltage source of twice the incident voltage wave. Figure 7(c) is Norton's equivalent circuit with the same impedance Z_o and a current source of twice the incident current wave. Either equivalent circuit can be used to calculate the terminal voltage.

Derivation of these circuits is simple. The voltage source in Theven-

[17] H. Amemiya, "Time-Domain Analysis of Multiple Parallel Transmission Lines by Means of Equivalent Circuits," *Electronics Letters*, Vol. 3, No. 1, p. 14, Jan. 1967.

in's equivalent circuit is the open-circuit voltage. This voltage is obtained by extending the transmission line to infinity, and by having two identical waves propagate toward each other as shown in Figure 8(a). Where the two waves meet, the net current is zero, corresponding to an open circuit. The voltages add, however, resulting in $2V$ for the open-circuit voltage. Similarly, the current source in Norton's

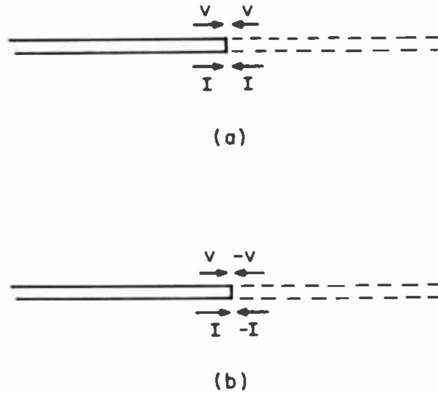


Fig. 8—Voltage source and current source in the equivalent circuits can be obtained by superposing two waves propagating in opposing directions: (a) Thevenin's voltage source (open-circuit voltage), (b) Norton's current source (short-circuit current).

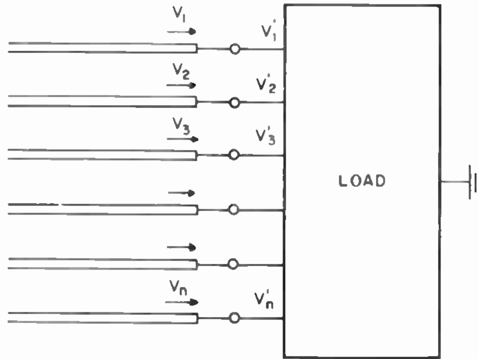
equivalent circuit is the short-circuit current. This current is $2I$ (Figure 8(b)) because the two waves are identical in amplitude but opposite in polarity.

The reflected wave can be obtaind as the difference between the incident wave and the terminal voltage (transmitted wave), $V' - V$. The reflection coefficient is defined as

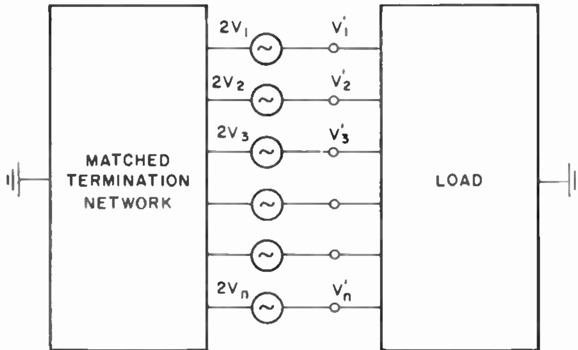
$$\rho = \frac{\text{reflected wave}}{\text{incident wave}} = \frac{V' - V}{V} = \frac{Z - Z_o}{Z + Z_o}.$$

The reflection coefficient is a function of the characteristic impedance and the load impedance, and is independent of the amplitude of the incident wave. Therefore, it is normal practice to calculate the reflection coefficient first, then to calculate the reflected wave and the terminal voltage. (Transmission coefficient $(1 + \rho)$ is used occasionally.) Because of the simplicity of the reflection coefficient approach, the equivalent circuits are not used frequently for these calculations for single transmission lines.

The equivalent-circuit concept can be applied to multiple parallel transmission lines. Transverse electric and magnetic (TEM) propagation is assumed, as it was for single transmission lines. In Figure 9(a), a system of n transmission lines and a ground with incident voltage



(a)



(b)

Fig. 9—Parallel transmission lines and equivalent circuit: (a) parallel transmission lines terminated into a load, (b) Thevenin's equivalent circuit replacing the transmission lines.

waves $V_1, V_2, \ldots, V_n$ is terminated with an arbitrary load. $V'_1, V'_2, \ldots, V'_n$ are terminal voltages and can be obtained from Thevenin's equivalent circuit in Figure 9(b). Terminal currents can also be obtained. The equivalent circuit consists of a matched termination network and source voltages with amplitudes twice those of the incident voltage waves. When terminating the transmission lines, the matched termination network absorbs all possible propagation modes without

reflections; i.e., for terminal conditions the matched-termination network is equivalent to parallel transmission lines that extend to infinity.

Norton's equivalent circuit can be used to represent multiple parallel transmission lines. This circuit consists of a matched termination network and source currents with amplitudes twice those of the incident current waves in a configuration similar to Figure 7(c) for a single-line case. These equivalent circuits can be derived using the same approach as for the single line.

The reflected waves can be obtained from the incident waves V_k and the terminal voltages V_k' as $V_k' - V_k$, where $k = 1, 2, \ldots, n$. Although a reflection coefficient can be defined for each line as the ratio of the reflected wave to the incident wave, its usefulness is questionable because it depends on the relative amplitudes of the incident waves.

As stated previously, a system of n parallel transmission lines and a ground has n propagation modes, each characterized by its own propagation constant. In one extreme, all propagation constants may be identical, a condition that occurs when the medium is homogeneous and isotropic. In the other extreme, all propagation constants may be different. However, the form of the equivalent circuit is not affected by any of these conditions.

Time-Domain Analyses of Multiple Parallel Transmission Lines by Equivalent Circuits

One useful application of equivalent circuits for multiple parallel transmission lines is in the time-domain analysis of wave propagation over such lines. Signal cross-talk analysis on printed-circuit transmission lines in high-speed computers is a typical example.

Using equivalent circuits, wave propagation on multiple parallel transmission lines can be studied according to the following steps:

1. Calculation of the sending-end voltages by substituting the matched-termination network for the transmission lines.
2. Separation of the sending-end voltages into mode components so that each propagation mode can be treated separately.
3. Modification of the voltage components in each mode according to the particular propagation constant and the line length.
4. Combination of the modified voltage components to obtain the incident voltage waves at the receiving end.
5. Calculation of the terminal voltages and the reflected waves by using Thevenin's equivalent circuit.

Step 2 is necessary because propagation constants are generally different from mode to mode. In each mode, the voltage amplitudes on

the lines have ratios specified by the eigenvector for the mode. The eigenvectors are obtained when solving the differential equations governing the transmission lines (see Appendix I). Using the eigenvectors, an arbitrary set of input voltages can be divided into sets of mode components, with each set appropriate for a particular propagation mode. When the propagation modes that are excited have identical propagation constants, it is not necessary to separate voltages into components, and a considerable simplification results. For lossless transmission lines, this condition implies the same propagation velocity for all propagation modes.

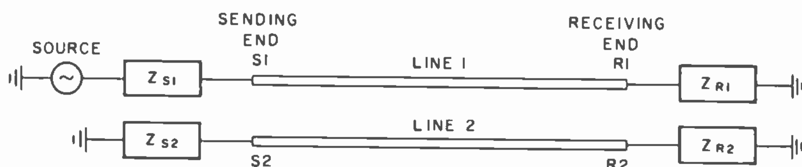


Fig. 10—Two parallel transmission lines.

For reflected waves, steps 2, 3, 4, and 5 are repeated. The receiving end is the new sending end. When the line parameters change with frequency, each frequency component in the source waveform must be subjected to the procedure separately.

Let us examine two lossless parallel transmission lines (Figure 10) to demonstrate the effectiveness of this approach. This structure is well known as a transmission-line directional coupler.[6-10] The directional-coupling action implies that when sending end S1 is driven externally, receiving end R2 will have no output if (1) the product of Z_{S2} and Z_{R1} is made equal to a certain constant and (2) this condition is not affected by Z_{S1} or by Z_{R2}. Papers on transmission-line directional couplers have dealt primarily with the frequency-domain behavior. Time-domain studies of the structure have largely assumed a loose coupling between the two lines and resistive termination of particular values.[10-15] With the procedure described here, a general case can be solved with no restriction on the degree of coupling. With resistive terminations of arbitrary values, only algebraic manipulation is required. Laplace transforms are necessary only when one or more of the termination impedances are reactive.

If we assume the same geometry for the two lines, the matched-termination network is as shown in Figures 6(c) and 6(d). Directional coupling action occurs when the common mode and the differential mode have the same propagation velocity.

Figure 11 shows the terminal voltage waveforms of the directional

coupler of Figure 10. These waveforms are obtained by following the five-step procedure described previously (see Appendix II for proof). The source is a step-function voltage and the termination impedances

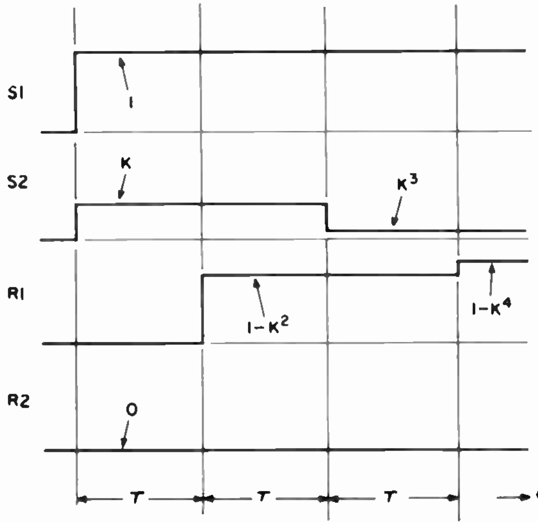


Fig. 11—Terminal voltage waveforms of the transmission line directional coupler of Figure 10; $Z_{S2} = Z_{R1} = \sqrt{Z_{oe} Z_{od}}$, K is the initial ratio of the sending-end terminal voltages, τ is the delay.

are chosen as follows:

Z_{S1} : resistive

$Z_{S2} = Z_{R1} = \sqrt{Z_{oe} Z_{od}}$ (resistive)

Z_{R2} : arbitrary

In this example, the directional coupling action occurs in both directions of propagation, and, as a result, the input impedance at sending end S1 is resistive and equal to $\sqrt{Z_{oe} Z_{od}}$. The waveforms are normalized, K indicates the initial voltage ratio at the sending end, and τ is the delay between the sending end and the receiving end. When K is small, the waveform at R1 is essentially the same as the waveform at S1 but delayed by τ. The waveform at S2 is a pulse of amplitude K and width 2τ. With an arbitrary source waveform, the waveform at S2 is,[11-15]

Waveform at S2 = K ((waveform at S1) − (waveform at S1 delayed by 2τ)).

When the two propagation modes have different delays, the directional coupling action is no longer possible. Voltage waves must be separated into common-mode components and differential-mode components. Figure 12 shows the terminal voltage waveforms of the

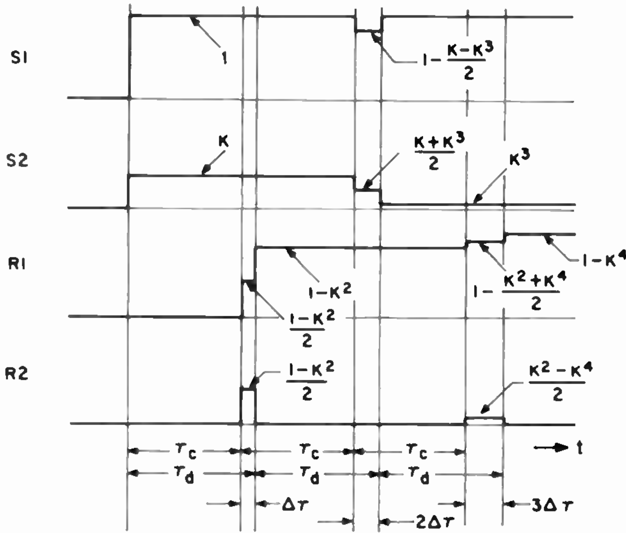


Fig. 12—Terminal voltage waveforms of the two coupled lines of Figure 10 with unequal delays for the two propagation modes. $Z_{S1} = Z_{S2} = Z_{R1} = Z_{R2} = \sqrt{Z_{oc} Z_{od}}$. K is the initial ratio of the sending-end terminal voltage. τ_c is the common-mode delay and τ_d is the differential-mode delay. $\Delta\tau = \tau_d - \tau_c$ is the delay difference. $\Delta\tau > 0$ in the figure.

coupled lines in which the common-mode delay is less than the differential-mode delay (see Appendix III for proof). The source is a step-function voltage and the terminations are

$$Z_{S1} = Z_{S2} = Z_{R1} = Z_{R2} = \sqrt{Z_{oc} Z_{od}}\,.$$

When the two delays are equal, the waveforms of Figures 11 and 12 are the same. When the common-mode delay is larger than the differential-mode delay, the notch in the sending-end waveform at S1 becomes a bump, and the positive pulses in the receiving-end waveform at R2 become negative pulses. The other two waveforms do not change.

Normally, K is small and the source waveform has rise and fall times much larger than the delay difference. Under these conditions, the waveform at R2 can be shown to be proportional to the line length and

to the time derivative of the source waveform.[13, 15] The waveform at S2 is given by the same formula used for directional coupler. If $K = 0$ (occurs when $Z_{oc} = Z_{od}$), a different type of directional coupler (a co-directional coupler) is obtained.[6] Sending end S2 (Figure 12) will now have no voltage. (Receiving end R2 has no voltage with a normal directional coupler.)

When the two lines have different geometries, the structure of the matched-termination network will be unbalanced. However, the same approach can be used to show the directional coupling action and the cross talk.

Acknowledgment

The author wishes to thank R. L. Pryor for the stimulating discussions which contributed to this paper.

Appendix I—Matched Termination Network for Multiple Parallel Transmission Lines

Let us consider a system of n parallel transmission lines and a ground. The following assumptions are made: (1) lines are lossless and (2) propagation modes are transverse electric and magnetic (TEM). Using matrix notation, the pertinent differential equations are:[3, 18]

$$\left[-\frac{dV}{dx} \right] = j\omega [L][I] \tag{2}$$

$$\left[-\frac{dI}{dx} \right] = j\omega [C][V] \tag{3}$$

$$[V] = \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ \cdot \\ \cdot \\ \cdot \\ V_n \end{bmatrix} \qquad [I] = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ \cdot \\ \cdot \\ \cdot \\ I_n \end{bmatrix}$$

[18] L. A. Pipes, *Matrix Methods for Engineering*, Prentice-Hall, Englewood Cliffs, N. J., 1963.

where

V_i = Voltage on line i

I_i = Current on line i

$i = 1, 2, 3, \ldots, n$

x = Distance in the direction of the lines

$$[L] = \begin{bmatrix} L_{11} & L_{12} & L_{13} & \ldots & L_{1n} \\ L_{21} & L_{22} & L_{23} & \ldots & L_{2n} \\ L_{31} & L_{32} & L_{33} & \ldots & L_{3n} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ L_{n1} & L_{n2} & L_{n3} & \ldots & L_{nn} \end{bmatrix}$$

where

L_{ii} = Self inductance per unit length of line i, $i = 1, 2, 3, \ldots, n$.

L_{ij} = Mutual inductance per unit length between line i and line j, $i \neq j$ and $i, j = 1, 2, 3, \ldots, n$.

$L_{ij} = L_{ji}$, $i, j = 1, 2, 3, \ldots, n$.

$$[C] = \begin{bmatrix} \sum_{i=1}^{n} C_{1i} & -C_{12} & -C_{13} & \ldots & -C_{1n} \\ -C_{21} & \sum_{i=1}^{n} C_{2i} & -C_{23} & \ldots & -C_{2n} \\ -C_{31} & -C_{32} & \sum_{i=1}^{n} C_{3i} & \ldots & -C_{3n} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ -C_{n1} & -C_{n2} & -C_{n3} & \ldots & \sum_{i=1}^{n} C_{ni} \end{bmatrix}$$

where

C_{ii} = Capacitance per unit length between line i and ground, $i = 1, 2, 3, \ldots, n$

C_{ij} = Capacitance per unit length between line i and line j, $i \neq j$ and $i, j = 1, 2, 3, \ldots, n$

$C_{ij} = C_{ji}$, $i, j = 1, 2, 3, \ldots, n$.

From Equations (2) and (3),

$$\left[\frac{d^2V}{dx^2}\right] = -\omega^2 [L][C][V] \tag{4}$$

Assume a solution of the form

$V_i = V_{oi}\, e^{j\beta x}$

$i = 1, 2, 3, \ldots, n$.

By substitution, Equation (4) becomes

$$\{\delta [U] - [L][C]\}[V] = 0 \tag{5}$$

where $\delta = \beta^2/\omega^2$ and $[U]$ is a unit matrix.

For the solution to be practical, we eliminate the trivial solution ($V_{oi} = 0$, $i = 1, 2, 3, \ldots, n$). Then the following must hold:

$$|\delta [U] - [L][C]| = 0 \tag{6}$$

Since $[L][C]$ is an nth order matrix, n values of δ are generally obtained. These values are called eigenvalues of $[L][C]$.

Propagation velocity and phase constants are given respectively by

$$v = 1/\sqrt{\delta} \text{ and } \beta = \pm \omega/v.$$

The positive sign is for backward propagation in the negative x direction, and the negative sign is for forward propagation in the positive x direction. When a matched termination network is studied, only one direction of propagation should be considered. Let us choose the forward propagation. Corresponding to n values of δ, there exist n values of propagation velocity. Each value is associated with a propagation mode.

When the medium is homogeneous and isotropic, all propagation velocities coincide, and the eigenvalues are identical. This occurs when

$$[L][C] = \delta_o [U] = \frac{1}{v_o^2}[U]\,. \tag{7}$$

Then Equation (6) gives

$$\delta = \delta_o \qquad (n \text{ equal roots})$$

and

$$v = v_o = \frac{1}{\sqrt{\delta_o}} \qquad \text{(common velocity)} .$$

A general solution is given by

$$[V] = \exp\{-j\frac{\omega}{v_o}x\}[V_o], \tag{8}$$

where elements of the column matrix $[V_o]$ can be chosen arbitrarily. Substituting Equation (8) into Equation (2), we obtain

$$[I] = \frac{1}{j\omega}[L]^{-1}\left[-\frac{dV}{dx}\right] = \frac{1}{v_o}[L]^{-1}[V] .$$

Using Equation (7), we get

$$[I] = v_o[C][V]. \tag{9}$$

Let $[G]$ be the conductance matrix of a matched termination network. Then,

$$[I] = [G][V] \tag{10}$$

where

$$[G] = \begin{bmatrix} \sum_{i=1}^{n} \frac{1}{R_{1i}} & -\frac{1}{R_{12}} & -\frac{1}{R_{13}} & \cdots & -\frac{1}{R_{1n}} \\ -\frac{1}{R_{21}} & \sum_{i=1}^{n} \frac{1}{R_{2i}} & -\frac{1}{R_{23}} & \cdots & -\frac{1}{R_{2n}} \\ -\frac{1}{R_{31}} & -\frac{1}{R_{32}} & \sum_{i=1}^{n} \frac{1}{R_{3i}} & \cdots & -\frac{1}{R_{3n}} \\ \vdots & \vdots & \vdots & & \vdots \\ -\frac{1}{R_{n1}} & -\frac{1}{R_{n2}} & -\frac{1}{R_{n3}} & \cdots & \sum_{i=1}^{n} \frac{1}{R_{ni}} \end{bmatrix}$$

R_{ii} = Resistance between terminal i and ground, $i = 1, 2, 3, \ldots, n$

R_{ij} = Resistance between terminal i and terminal j, $i \neq j$ and i, $j = 1, 2, 3, \ldots, n$

$R_{ij} = R_{ji}$, $i, j = 1, 2, 3, \ldots, n$.

By comparing Equation (9) and Equation (10), we find $[G] = v_o[C]$. Therefore

$$R_{ij} = 1/(v_o C_{ij}), \quad i, j = 1, 2, 3, \ldots, n. \tag{11}$$

as given in the text.

When propagation velocities do not coincide, Equation (11) is unusable. Inhomogeneous media and artificial transmission lines generally produce this condition. It is possible to find the conductance matrix of a matched termination network, but the process is much more involved. Assume that all propagation velocities are different. Then from Equation (6),

$\delta = \delta_i$ eigenvalues of $[L][C]$

$v_i = 1/\sqrt{\delta}$ propagation velocities

$i = 1, 2, 3, \ldots, n$.

For the ith propagation mode with δ_i and v_i, a set of n voltages, one for each of n lines, is determined by using Equation (5). This set is called an eigenvector. An eigenvector has a direction without specified amplitude, that is, only the ratios between the voltages are meaningful. A solution for the ith mode is then given by

$$[V]_i = \exp \left\{ -j \frac{\omega}{v_i} x \right\} [V_o]_i \tag{12}$$

$$[V]_i = \begin{bmatrix} V_{1i} \\ V_{2i} \\ V_{3i} \\ \cdot \\ \cdot \\ V_{ni} \end{bmatrix} \qquad [V_o]_i = \begin{bmatrix} V_{o1i} \\ V_{o2i} \\ V_{o3i} \\ \cdot \\ \cdot \\ V_{oni} \end{bmatrix}$$

$$i = 1, 2, 3, \ldots, n,$$

where both $[V]_i$ and $[V_o]_i$ are eigenvectors.

Substituting Equation (12) into Equation (2), we get

$$[I]_i = \frac{1}{v_i}[L]^{-1}[V]_i \tag{13}$$

$$[I]_i = \begin{bmatrix} I_{1i} \\ I_{2i} \\ I_{3i} \\ \cdot \\ \cdot \\ I_{ni} \end{bmatrix}$$

$$i = 1, 2, 3, \ldots, n.$$

$[V]_i$ and $[I]_i$ are column matrices, where i can be any one of the integers, 1 through n. It is possible to form square matrices whose columns are those of the column matrices with different i:

$$[V]_m = \begin{bmatrix} V_{11} & V_{12} & V_{13} & \cdots & V_{1n} \\ V_{21} & V_{22} & V_{23} & \cdots & V_{2n} \\ V_{31} & V_{32} & V_{33} & \cdots & V_{3n} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ V_{n1} & V_{n2} & V_{n3} & \cdots & V_{nn} \end{bmatrix}$$

$$[I]_m = \begin{bmatrix} I_{11} & I_{12} & I_{13} & \cdots & I_{1n} \\ I_{21} & I_{22} & I_{23} & \cdots & I_{2n} \\ I_{31} & I_{32} & I_{33} & \cdots & I_{3n} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ I_{n1} & I_{n2} & I_{n3} & \cdots & I_{nn} \end{bmatrix}$$

A diagonal matrix is formed by placing $\frac{1}{v}$'s on the diagonal.

$$\left[\frac{1}{v}\right] = \begin{bmatrix} \frac{1}{v_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{v_2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{v_3} & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & \frac{1}{v_n} \end{bmatrix}$$

Then, from Equation (13),

$$[I]_m = [L]^{-1} [V]_m \left[\frac{1}{v}\right]. \tag{14}$$

It is known from the matrix theory

$$[V]_m^{-1} [L][C][V]_m = [\delta] = \left[\frac{1}{v^2}\right], \tag{15}$$

where

$$\left[\frac{1}{v^2}\right] = \begin{bmatrix} \frac{1}{v_1^2} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{v_2^2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{v_3^2} & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & \frac{1}{v_n^2} \end{bmatrix}$$

Equation (15) can be rewritten as

$$[C][V]_m[v] = [L]^{-1}[V]_m\left[\frac{1}{v}\right] \tag{16}$$

where

$$[v] = \begin{bmatrix} v_1 & 0 & 0 & \dots & 0 \\ 0 & v_2 & 0 & \dots & 0 \\ 0 & 0 & v_3 & \dots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \dots & v_n \end{bmatrix}$$

Combining Equations (14) and (16), we have

$$[I]_m = [C][V]_m[v] = [L]^{-1}[V]_m\left[\frac{1}{v}\right]. \tag{17}$$

The conductance matrix of a matched termination network must satisfy

$$[I]_m = [G][V]_m. \tag{18}$$

From Equations (17) and (18),

$$[G] = [C][V]_m[v][V]_m^{-1} = [L]^{-1}[V]_m\left[\frac{1}{v}\right][V]_m^{-1}. \tag{19}$$

By Equation (19), a matched termination network can be computed from inductance and capacitance parameters of the given transmission lines. $[V]_m$, which is sometimes called a modal matrix, is made of n different eigenvectors of $[L][C]$. Therefore $[V]_m$ is not uniquely defined. However $[G]$ is uniquely determined by Equation (19).

For a matched termination network to be physically realizable, $[G]$ must be symmetric. From Equation (19),

$$[C]^{-1}[G] = [V]_m[v][V]_m^{-1}$$

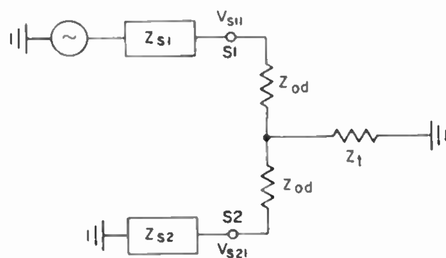
and

$$[L][G] = [V]_m\left[\frac{1}{v}\right][V]_m^{-1}.$$

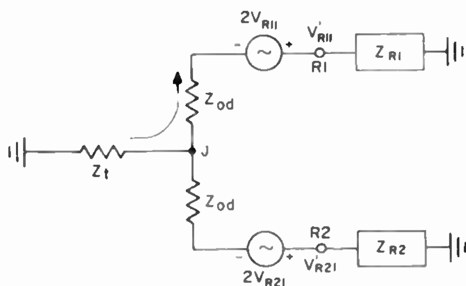
or

$$-\frac{2V_{R11}}{Z_{od}+Z_t+Z_{R1}}Z_t = -2V_{R21}.$$

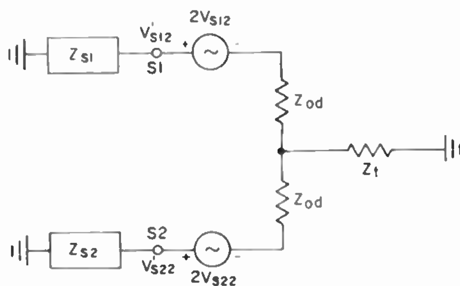
The implication of the preceding equation is that due to the current produced by $2V_{R11}$, the potential of point J of Figure 14(b) is equal but



(a)



(b)



(c)

Fig. 14—Equivalent circuits for the directional coupler of Figure 10.

opposite in sign to $2V_{R21}$, thus making V_{R21}' zero. This is valid even though point J is a fictitious point.

Now the functions of Z_{S2} and Z_{R1} in the directional coupling action are clear. Z_{S2} controls the ratio between the primary waves V_{S21} and V_{S11} and, consequently, the ratio between the incident waves V_{R21} and V_{R11} at the receiving end. Z_{R1} controls the reaction from V_{R11} through Z_t so that the terminal voltage V_{R21}' is zero. For instance, when Z_{S2} is zero, V_{S21} and V_{R21} are zero, which means no reaction is desired. Therefore, Z_{R1} is made an open circuit.

It is necessary to consider reflected waves to prove the directional coupling action conclusively. The reflected waves at the receiving end are:

$$V_{R12} = V_{R11}' - V_{R11} = -\frac{Z_{od} + Z_t - Z_{R1}}{Z_{od} + Z_t + Z_{R1}} V_{R11},$$

$$V_{R22} = V_{R21}' - V_{R21} = -V_{R21} = -\frac{Z_t}{Z_{od} + Z_t + Z_{R1}} V_{R11},$$

$$\frac{V_{R12}}{V_{R22}} = \frac{Z_{od} + Z_t - Z_{R1}}{Z_t}.$$

V_{R12} and V_{R22} propagate toward the sending end and become incident waves V_{S12} and V_{S22} at the sending end. V_{S12} and V_{S22} are identical to V_{R12} and V_{R22}, respectively, except for the line delay. Therefore,

$$\frac{V_{S12}}{V_{S22}} = \frac{V_{R12}}{V_{R22}} = \frac{Z_{od} + Z_t - Z_{R1}}{Z_t}.$$

The terminal voltages V_{S12}' and V_{S22}' and the reflected waves V_{S13} and V_{S23} can be calculated by using the equivalent circuit of Figure 14(c). The calculation, which requires some manipulation, shows that Z_{S1} does not affect the ratio between V_{S23} and V_{S13} although it does affect individual amplitudes;

$$\frac{V_{S23}}{V_{S13}} = \frac{Z_t}{Z_{od} + Z_t + Z_{R1}}.$$

This is the ratio needed between the two primary waves to obtain the directional coupling action. Therefore, the tertiary waves will produce no output at receiving end $R2$. It can be proved by induction that the two waves on the two lines propagating toward the receiving end always have the specific ratio that produces the directional coupling action;

$$\frac{V_{S21}}{V_{S11}} = \frac{V_{S23}}{V_{S13}} = \frac{V_{S25}}{V_{S15}} = \ldots = \frac{Z_t}{Z_{od} + Z_t + Z_{R1}}.$$

It can also be proved that the two waves on the two lines propagating toward the sending end have a constant ratio;

$$\frac{V_{R12}}{V_{R22}} = \frac{V_{R14}}{V_{R24}} = \ldots = \frac{Z_{od} + Z_t - Z_{R1}}{Z_t}.$$

This ratio is normally not proper to obtain the directional coupling action. However, if

$$Z_{S2} = Z_{R1} = \sqrt{Z_{oe} Z_{od}},$$

then

$$\frac{V_{S21}}{V_{S11}} = \frac{V_{R12}}{V_{R22}} = \frac{V_{S23}}{V_{S13}} = \frac{V_{R14}}{V_{R24}} = \frac{V_{S25}}{V_{S15}} = \ldots$$

$$= \frac{Z_t}{Z_{od} + Z_t + \sqrt{Z_{oe} Z_{od}}} = \frac{Z_{od} + Z_t - \sqrt{Z_{oe} Z_{od}}}{Z_t}$$

$$= \frac{Z_{oe} - Z_{od}}{Z_{oe} + Z_{od} + 2\sqrt{Z_{oe} Z_{od}}} = \frac{Z_{oe} + Z_{od} + 2\sqrt{Z_{oe} Z_{od}}}{Z_{oe} - Z_{od}} = K,$$

which implies that the directional coupling action occurs in both directions of propagation. In other words, not only is the terminal voltage at receiving end R2 zero, but the terminal voltage at sending end S1 is also zero for all higher-order waves. Therefore, the input impedance at terminal S1 becomes resistive when the preceding condition is satisfied. The input impedance is determined from Figure 14(a) to be $\sqrt{Z_{oe} Z_{od}}$, for $Z_{S2} = \sqrt{Z_{oe} Z_{od}}$. When a terminal voltage is zero, the

reflected wave is the same as the inverted incident wave; therefore the relative wave amplitudes can be easily calculated. In the following equations, the waves are expressed as a function of time, $V_{S11}(t)$, and the delays are indicated (τ is the one-way delay);

$$
\begin{aligned}
&V_{S21}(t) = KV_{S11}(t),\\
&V_{R11}(t) = V_{S11}(t-\tau),\\
&V_{R21}(t) = KV_{S11}(t-\tau),\\
&V_{R22}(t) = -V_{R21}(t) = -KV_{S11}(t-\tau),\\
&V_{R12}(t) = KV_{R22}(t) = -K^2V_{S11}(t-\tau),\\
&V_{S22}(t) = -KV_{S11}(t-2\tau),\\
&V_{S12}(t) = -K^2V_{S11}(t-2\tau),\\
&V_{S13}(t) = -V_{S12}(t) = K^2V_{S11}(t-2\tau),\\
&V_{S23}(t) = KV_{S13}(t) = K^3V_{S11}(t-2\tau),\\
&V_{R13}(t) = K^2V_{S11}(t-3\tau),\\
&V_{R23}(t) = K^3V_{S11}(t-3\tau),\\
&V_{R24}(t) = -V_{R23}(t) = -K^3V_{S11}(t-3\tau),\\
&V_{R14}(t) = KV_{R24}(t) = -K^4V_{S11}(t-3\tau),
\end{aligned}
$$

and so on.

From these wave equations the voltages at the four terminals can be obtained.

At sending end S1,

$$
\begin{aligned}
V_{S1}(t) &= V_{S11}(t) + V_{S12}(t) + V_{S13}(t) + \dots\\
&= V_{S11}(t) - K^2V_{S11}(t-2\tau) + K^2V_{S11}(t-2\tau) + \dots\\
&= V_{S11}(t)\,.
\end{aligned}
$$

At sending end S2,

$$
\begin{aligned}
V_{S2}(t) &= V_{S21}(t) + V_{S22}(t) + V_{S23}(t) + \dots\\
&= KV_{S11}(t) - KV_{S11}(t-2\tau) + K^3V_{S11}(t-2\tau) - \dots\,.
\end{aligned}
$$

At receiving end R1,

$$
\begin{aligned}
V_{R1}(t) &= V_{R11}(t) + V_{R12}(t) + V_{R13}(t) + V_{R14}(t) + \dots\\
&= V_{S11}(t-\tau) - K^2V_{S11}(t-\tau) + K^2V_{S11}(t-3\tau)\\
&\qquad - K^4V_{S11}(t-3\tau) + \dots\,.
\end{aligned}
$$

At receiving end R2,

$$
\begin{aligned}
V_{R2}(t) &= V_{R21}(t) + V_{R22}(t) + V_{R23}(t) + V_{R24}(t) + \dots\\
&= KV_{S11}(t-\tau) - KV_{S11}(t-\tau) + K^3V_{S11}(t-3\tau)\\
&\qquad - K^3V_{S11}(t-3\tau) + \dots = 0\,.
\end{aligned}
$$

Waveforms in Figure 11 are obtained from the previous equations by assuming a step function for the source voltage.

In the preceding discussion of directional coupling action, Z_{S2} and Z_{R1} were considered resistive. This is not a necessary condition, however, and a similar approach can be used to prove directional coupling action by replacing the voltages with their Laplace transforms. Z_{S2} and Z_{R1} may be reactive, but the product must be real, because Z_{oc} and Z_{od} are both resistive; i.e., when either Z_{S2} or Z_{R1} is capacitive, the other must be inductive. Also, equal geometries were assumed for the two transmission lines. Accordingly, the matched termination network has a balanced structure with two equal arms as shown in Figure 6(c) and (d). When the lines have unequal geometries, these two arms will become different, but otherwise the same derivation will hold.

To obtain directional coupling action, the common mode and the differential mode must have the same propagation velocity. Otherwise, the cancellation effect which made a receiving-end terminal voltage zero is no longer possible because of the waveform distortion caused by the two components propagating with different velocities.

Appendix III—Crosstalk

When the common mode and the differential mode have different delays, the directional coupling action ceases. As noted in the text, waves must be separated into common-mode and differential-mode components. After each one-way propagation, they are combined to obtain the incident waves. However, the paticular terminations, $Z_{S1} = Z_{S2} = Z_{R1} = Z_{R2} = \sqrt{Z_{oc} Z_{od}}$, used for Figure 12, permit some simplification. When the two balanced lines are terminated into two equal impedances, there is no mode conversion. The common-mode incident waves will produce common-mode reflected waves only, and similarly the differential-mode incident waves will produce differential-mode reflected waves only. Therefore, combination of the two components is not required after each propagation. Of course, to obtain the true waveforms at the terminals, all components involved must be combined.

The voltage ratio K is defined in the same manner as previously:

$$\frac{V_{S21}(t)}{V_{S11}(t)} = \frac{Z_t}{Z_{od} + Z_t + \sqrt{Z_{oc} Z_{od}}} = K .$$

A common-mode component and a differential-mode component is defined for each pair of waves;

Sending end:

Common mode $$V_{SCJ}(t) = \frac{1}{2}\{V_{S1J}(t) + V_{S2J}(t)\},$$

Differential mode $$V_{SDJ}(t) = \frac{1}{2}\{V_{S1J}(t) - V_{S2J}(t)\},$$

$$J = 1, 2, 3, \ldots$$

Receiving end:

Common mode $$V_{RCJ}(t) = \frac{1}{2}\{V_{R1J}(t) + V_{R2J}(t)\},$$

Differential mode $$V_{RDJ}(t) = \frac{1}{2}\{V_{R1J}(t) - V_{R2J}(t)\},$$

$$J = 1, 2, 3, \ldots$$

The subscript for these components consists of two letters and one numeral (J indicates a numeral). The first letter, which is either S or R, indicates either the sending end or the receiving end, respectively. The second letter, which is either C or D, indicates either the common mode or the differential mode. The numeral indicates the sequence of waves.

The following relations hold:

Common Mode	*Differential Mode*
$V_{SC1}(t) = \frac{1+K}{2} V_{S11}(t)$	$V_{SD1}(t) = \frac{1-K}{2} V_{S11}(t)$
$V_{RC1}(t) = \frac{1+K}{2} V_{S11}(t-\tau_c)$	$V_{RD1}(t) = \frac{1-K}{2} V_{S11}(t-\tau_d)$
$V_{RC2}(t) = -\frac{(1+K)K}{2} V_{S11}(t-\tau_c)$	$V_{RD2}(t) = \frac{(1-K)K}{2} V_{S11}(t-\tau_d)$
$V_{SC2}(t) = -\frac{(1+K)K}{2} V_{S11}(t-2\tau_c)$	$V_{SD2}(t) = \frac{(1-K)K}{2} V_{S11}(t-2\tau_d)$

Common Mode

$$V_{SC3}(t) = \frac{(1+K)K^2}{2} V_{S11}(t - 2\tau_c)$$

$$V_{RC3}(t) = \frac{(1+K)K^2}{2} V_{S11}(t - 3\tau_c)$$

$$V_{RC4}(t) = -\frac{(1+K)K^3}{2} V_{S11}(t - 3\tau_c)$$

etc.,

where

τ_c = common-mode delay.

Differential Mode

$$V_{SD3}(t) = \frac{(1-K)K^2}{2} V_{S11}(t - 2\tau_d)$$

$$V_{RD3}(t) = \frac{(1-K)K^2}{2} V_{S11}(t - 3\tau_d)$$

$$V_{RD4}(t) = \frac{(1-K)K^3}{2} V_{S11}(t - 3\tau_d)$$

etc.,

where

τ_d = differential-mode delay.

The terminal voltages are calculated separately for each mode:

(a) *Common mode*

Sending end S1 and S2:

$$V_{SC}(t) = V_{SC1}(t) + V_{SC2}(t) + V_{SC3}(t) + \ldots$$

$$= \frac{1+K}{2} V_{S11}(t) - \frac{(1+K)K}{2} V_{S11}(t-2\tau_c) + \frac{(1+K)K^2}{2} V_{S11}(t-2\tau_c) + \ldots$$

$$= \frac{1+K}{2} V_{S11}(t) - \frac{(1-K^2)K}{2} V_{S11}(t-2\tau_c) - \ldots$$

Receiving end R1 and R2:

$$V_{RC}(t) = V_{RC1}(t) + V_{RC2}(t) + V_{RC3}(t) + V_{RC4}(t) + \ldots$$

$$= \frac{1+K}{2} V_{S11}(t-\tau_c) - \frac{(1+K)K}{2} V_{S11}(t-\tau_c)$$

$$+ \frac{(1+K)K^2}{2} V_{S11}(t-3\tau_c) - \frac{(1+K)K^3}{2} V_{S11}(t-3\tau_c) + \ldots$$

$$= \frac{1 - K^2}{2} V_{S11}(t - \tau_c) + \frac{(1 - K^2)K^2}{2} V_{S11}(t - 3\tau_c) + \dots .$$

(b) *Differential Mode*

Sending end S1 (change sign for sending end S2) :

$$V_{SD}(t) = V_{SD1}(t) + V_{SD2}(t) + V_{SD3}(t) + \dots$$

$$= \frac{1 - K}{2} V_{S11}(t) + \frac{(1 - K)K}{2} V_{S11}(t - 2\tau_d)$$

$$+ \frac{(1 - K)K^2}{2} V_{S11}(t - 2\tau_d) + \dots$$

$$= \frac{1 - K}{2} V_{S11}(t) + \frac{(1 - K^2)K}{2} V_{S11}(t - 2\tau_d) + \dots .$$

Receiving end R1 (change sign for receiving end R2) :

$$V_{RD} = V_{RD1}(t) + V_{RD2}(t) + V_{RD3}(t) + V_{RD4}(t) + \dots$$

$$= \frac{1 - K}{2} V_{S11}(t - \tau_d) + \frac{(1 - K)K}{2} V_{S11}(t - \tau_d)$$

$$+ \frac{(1 - K)K^2}{2} V_{S11}(t - 3\tau_d) + \frac{(1 - K)K^3}{2} V_{S11}(t - 3\tau_d) + \dots$$

$$= \frac{1 - K^2}{2} V_{S11}(t - \tau_d) + \frac{(1 - K^2)K^2}{2} V_{S11}(t - 3\tau_d) + \dots .$$

Figure 15 shows the common-mode and the differential-mode components of the terminal voltages when the source voltage is a step function. The true terminal voltages are obtained by combining the two components:

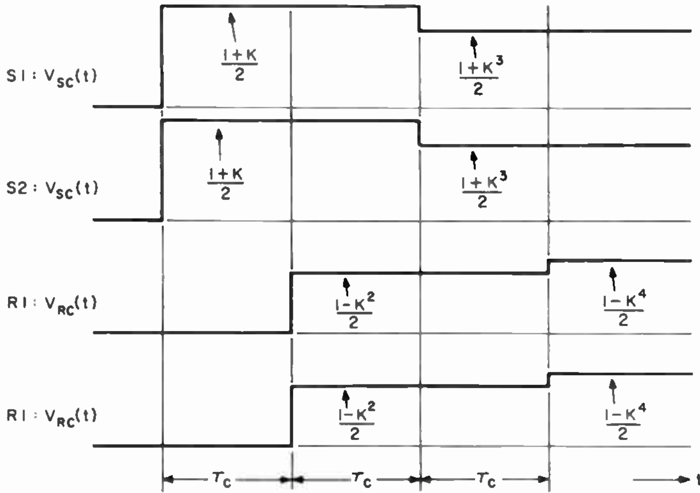
Sending end S1: $V_{S1}(t) = V_{SC}(t) + V_{SD}(t)$,

Sending end S2: $V_{S2}(t) = V_{SC}(t) - V_{SD}(t)$,

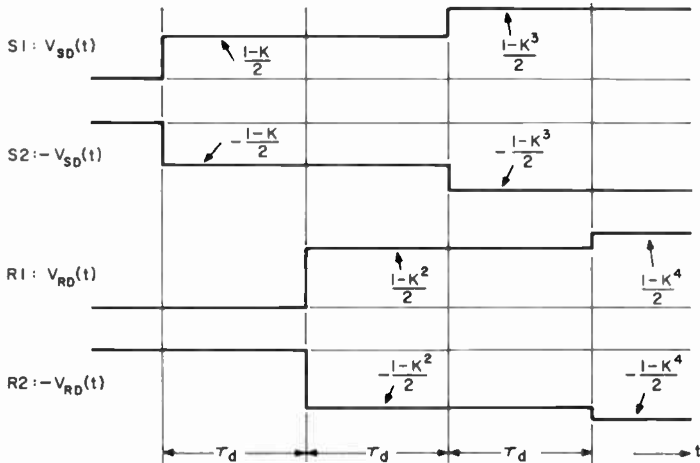
Receiving end R1: $V_{R1}(t) = V_{RC}(t) + V_{RD}(t)$,

Receiving end R2: $V_{R2}(t) = V_{RC}(t) - V_{RD}(t)$.

Thus, by adding the corresponding waveform components shown in Figure 15, the terminal voltage waveforms of Figure 12 are obtained.



(a) Common-mode component



(b) Differential-mode component

Fig. 15—Common-mode and differential-mode components of the terminal voltages; (a) common-mode component; (b) differential-mode component.

UNLOCK BEHAVIOR OF THE SECOND-ORDER PHASE-LOCKED LOOP WITH AND WITHOUT INTERFERING CARRIERS

BY

F. S. KEBLAWI

RCA Astro-Electronics Division
Princeton, N. J.

Summary—The mean time to unlock of a second-order phase-locked loop was experimentally measured for a single carrier perturbed by white gaussian noise and for the same carrier perturbed by both white gaussian noise and another interfering carrier. The latter differed from the tracked carrier by both frequency and amplitude. The experimental technique used is based on correlating the output of a strongly driven loop with that of a loop operating near threshold. Experimental results agreed reasonably well with available theoretical results.

INTRODUCTION

SEVERAL theoretical investigations of the behavior of second-order loop perturbed by noise have been published.[1-3] Sanneman and Rowbotham[1] presented results of a digital simulation of the loop. Their results relate, in effect, the normalized mean time of unlock (NMTU) and the signal-to-noise ratio in the loop bandwidth, $(S/N)_{2B_L}$. They encompass the range

$$5 < \text{NMTU} < 600$$

$$-2 \text{ db} < (S/N)_{2B_L} < 6 \text{ db} ,$$

Viterbi's[2] results are basically for a first-order loop. Charles and Lindsey[3] presented distributions of the intervals between cycle-skipping events but did not present the mean time to unlock at different signal-to-noise ratios. However, this information can be obtained from the distributions they present.

[1] R. W. Sanneman and J. R. Rowbotham, "Unlock Characteristics of the Optimum Type II Phase Locked Loop," *IEEE Trans. on Aerospace and Navigational Electronics*, March 1964.

[2] A. J. Viterbi, "Phase Locked Dynamics in the Presence of Noise by Fokker-Planck Techniques," The University of Michigan Engineering Summer Conference on Communication Theory, July 1965.

[3] F. J. Charles and W. C. Lindsey, "Some Analytical and Experimental Phase Locked Loop Results for Low Signal to Noise Ratios." JPL Contract No. NAS-7-100.

Thus, to the best of the author's knowledge, there is no direct experimental evaluation of the mean time to unlock as a function of the signal-to-noise ratio. Furthermore, no analyses or data could be found to assist in evaluating the limitations of coherent types of modulation systems in a multipath environment. This was of importance for a planetary exploration capsule relay link system design and it is in support of this system design that the present study has been undertaken.

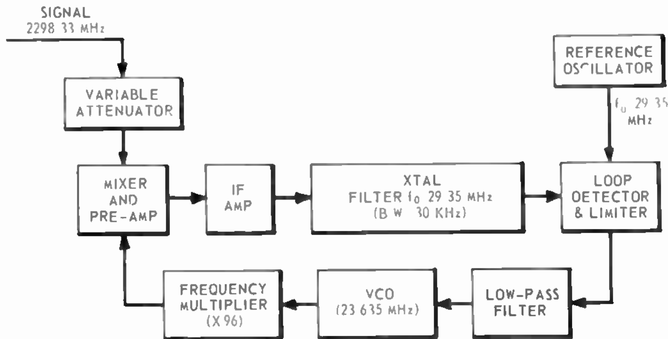


Fig. 1—Block diagram of test receiver.

Information was obtained on the unlock behavior of the loop for an interfering carrier in addition to thermal noise perturbing the tracked carrier. Unfortunately, no theoretical analysis was available with which these results could be compared. However, the only meaningful comparison of the loop behavior with an interfering carrier and thermal noise is with the loop behavior perturbed by thermal noise alone.

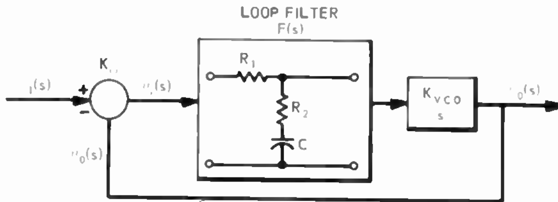
Test Receiver

Figure 1 is a block diagram of the test receiver used. The tracked r-f input-signal frequency fed into the receiver is 2298.33 MHz. The loop voltage-controlled oscillator (VCO) has a center frequency of 23.635 MHz. The VCO frequency is multiplied by 96 and then mixed with the input signal to the receiver. Some amplification is provided in the i-f stage of the receiver in addition to the preamplification following the mixing. The signal is then passed through a crystal filter having a center frequency of 29.35 MHz and a bandwidth of 30 kHz. Following band limiting, the signal is amplitude limited prior to phase detection, which is accomplished by multiplying the signal by the output of a reference oscillator whose frequency is 29.35 MHz. The output

of the phase detector is then filtered by the loop low-pass filter. The output of the latter drives the loop VCO.

NOISE BANDWIDTH OF THE LOOP

The linearized transfer function of the second-order loop (see Figure 2) is given by



$$F(s) = \frac{T_2 s + 1}{T_1 s + 1} \qquad \omega_n^2 = \frac{K}{T_1}$$

$$T_1 = (R_1 + R_2)C \qquad K = K_\phi K_{vco} = \text{static gain of the loop}$$

$$T_2 = R_2 C \qquad \zeta = \frac{T_2 \omega_n}{2}$$

Fig. 2—The second-order phase-locked-loop diagram and related parameters.

$$H(s) = \frac{\theta_o(s)}{\theta_i(s)} = \frac{2\,\zeta\,\omega_n\,s + \omega_n^2}{s^2 + 2\,\zeta\,\omega_n s + \omega_n^2}, \tag{1}$$

where θ_o, θ_i = phase of output and input signals, respectively, in degrees (or radians)

ζ = damping factor (dimensionless)

ω_n = undamped natural frequency of the loop (radians/sec)

$H(s)$ describes the frequency response of the loop for any ζ. Due to the variation of ζ with the signal strength (or the gain of the system), the frequency response varies with the signal level. Another factor that affects the frequency response is the limiter preceding the loop.

The frequency response of the loop was found experimentally by frequency modulating the input signal to the loop and taking the ratio of the signal amplitude at the loop VCO input (see Figure 1) to that of the modulating signal. This ratio as a function of the modulating frequency is the closed-loop frequency response to within a multiplicative constant. Figures 3, 4, and 5 show the frequency responses of the

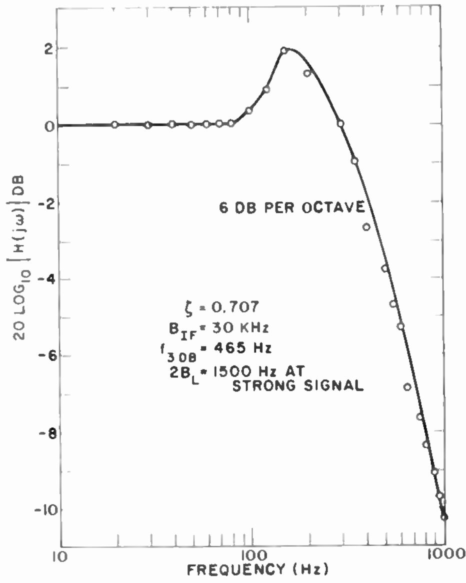


Fig. 3—Closed-loop frequency response for strong signal.

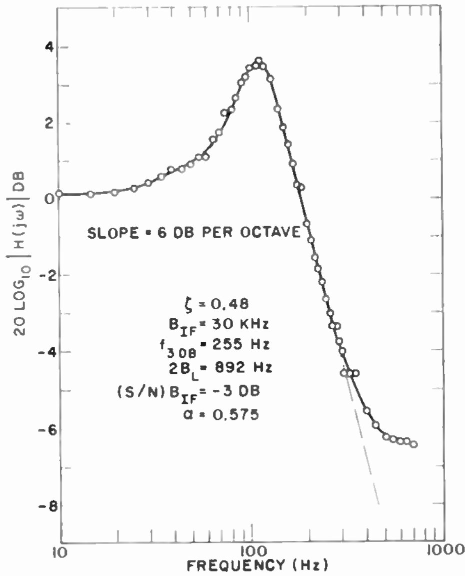


Fig. 4—Closed-loop frequency response.

loop for different signal levels. These figures illustrate that the frequency response varies with signal level when the latter is below a certain value. This is discussed further in a later section.

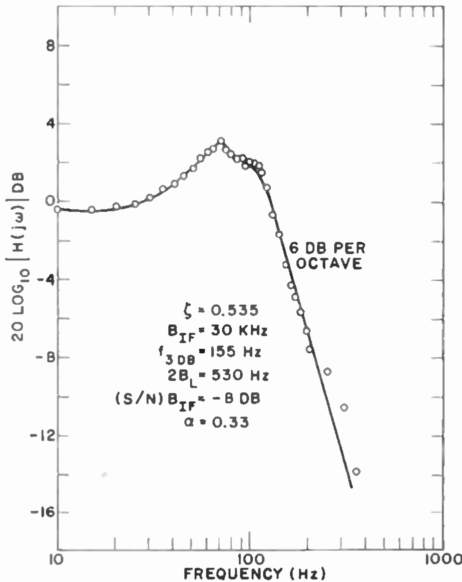


Fig. 5—Closed-loop frequency response.

Two parameters of the frequency-response curve are observed to vary with signal strength:

(1) The magnitude of the peak and the frequency at which it occurs.

(2) The 3-db bandwidth of the loop, which varied between 155 Hz and 465 Hz over the measured range of signal-to-noise ratio. Thus, the 3-db bandwidth continues to increase with increased signal strength up to the point where the suppression factor of the limiter is unity.[6,7] The loop noise bandwidth also varies with the variation of the 3-db bandwidth, although not linearly. This variation is described by the following relations[4,5]

---

[4] T. J. Karras, "Equivalent Noise Bandwidth Analysis from Transfer Functions," NASA Technical Note TND-2842, Goddard Space Flight Center, Greenbelt, Maryland.

[5] F. M. Gardner, S. S. Kent, and R. D. Dasenbrock, "Theory of Phase Lock Techniques," NASA N-66-10515.

[6] B. D. Martin, "The Pioneer IV Lunar Probe: A Minimum Power FM/PM System Design," JPL TR 32-215 March 15, 1962.

[7] W. B. Davenport, Jr., "Signal to Noise Ratios in Bandpass Limiters," *Jour. Appl. Phys.*, Vol. 24, p. 270, June 1953.

$$2B_L = \left(\frac{4\,\zeta^2 + 1}{4\,\zeta}\right)\omega_n \tag{2}$$

$$\left[\frac{\omega_{3db}}{\omega_n}\right]^2 = 1 + 2\,\zeta^2 + [(1 + 2\,\zeta^2)^2 + 1]^{1/2}, \tag{3}$$

or

$$2B_L = \frac{4\,\zeta^2 + 1}{4\,\zeta} \times \frac{\omega_{3db}}{(1 + 2\,\zeta^2 + [(1 + 2\,\zeta^2)^2 + 1]^{1/2})^{1/2}}. \tag{4}$$

where $2B_L$ is the double-sided loop noise bandwidth and ω_{3db} is the 3-db bandwidth of the loop. In order to obtain the signal-to-noise ratio in the loop bandwidth, $(S/N)_{2B_L}$, for any S/N ratio in the i-f bandwidth, $(S/N)_{B_{\mathrm{if}}}$, i.e., at the output of the crystal filter, it is necessary to take into account the variation of the loop noise bandwidth with signal level. An illustration of $(S/N)_{2B_L}$ versus $(S/N)_{B_{\mathrm{if}}}$ is presented later. Here, the experimental evaluation of the loop noise bandwidth is considered.

With Equation (4), the loop noise bandwidth can be obtained from a knowledge of the 3-db bandwidth and ζ, both of which can be obtained experimentally from the frequency response curve. While the 3-db bandwidth is obtained directly, the damping factor can be obtained from a knowledge of the peak of the response curve as follows. In general, for any ζ,

$$|H(\omega)|^2 = \frac{\omega_n^2 + 4\,\zeta^2\,\omega_n^2\omega^2}{(\omega_n^2 - \omega^2)^2 + 4\,\zeta^2\,\omega_n^2\omega^2} = \frac{1 + 4\,\zeta^2\left(\dfrac{\omega}{\omega_n}\right)^2}{\left(1 - \dfrac{\omega^2}{\omega_n^2}\right)^2 + 4\,\zeta^2\left(\dfrac{\omega}{\omega_n}\right)^2}. \tag{5}$$

If we let $(\omega/\omega_n)^2 = \chi$ and $|H(\omega)|^2 = A(\chi)$, (6)

$$A(\chi) = \frac{1 + 4\,\zeta^2\chi}{(1-\chi)^2 + 4\,\zeta^2\,\chi}. \tag{7}$$

The peak of this function occurs at

$$\chi_{\mathrm{peak}} = \frac{1}{4\,\zeta^2}\,[\sqrt{1 + 8\,\zeta^2} - 1]. \tag{8}$$

For different damping factors, Equation (8) can be substituted into Equation (7) and the $[A(\chi)]_{peak}$ is found in db's above the 0-db level, which corresponds to the low-frequency response. $|H(\omega)|_{peak}$ is plotted as a function of ζ in Figure 6. Figure 7 is a plot of Equation (8). Thus, an experimental knowledge of peak of the frequency response yields ζ by the use of Figure 6.

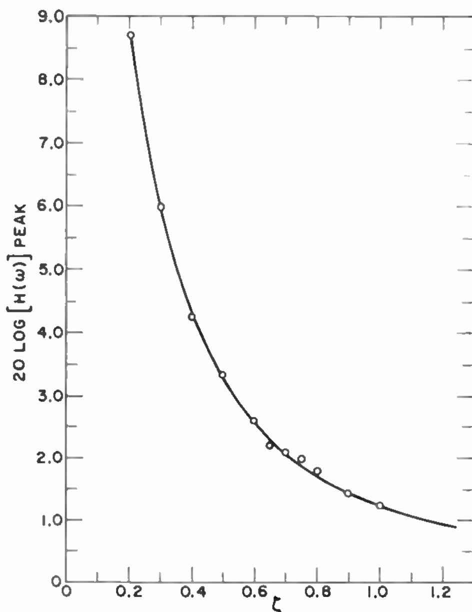


Fig. 6—$|H(\omega)|_{peak}$ versus ζ.

The noise bandwidths for Figures 3, 4, and 5 were calculated using Equation (4), where ζ was determined by correlating the peaks of the frequency response curves of these figures with Figure 7.

Limiter Suppression Factor

The limiter suppression factor, α, as given in References (6) and (7), is shown in Figure 8. These references give the following relation for small $(S/N)_i = (S/N)_{B_{if}}$ (the signal-to-noise ratio at the input to the limiter),

$$\alpha^2 \approx \frac{1}{1 + \frac{4}{\pi}\left(\frac{S}{N}\right)_i}. \tag{9}$$

For $(S/N)_{B_{if}} < -20$ db, this can be simplified to

$$\alpha^2 \approx \frac{\pi}{4}\left(\frac{S}{N}\right)_i . \tag{10}$$

Equation (10) is plotted in Figure 9.

The suppression factor at threshold can be obtained for any loop

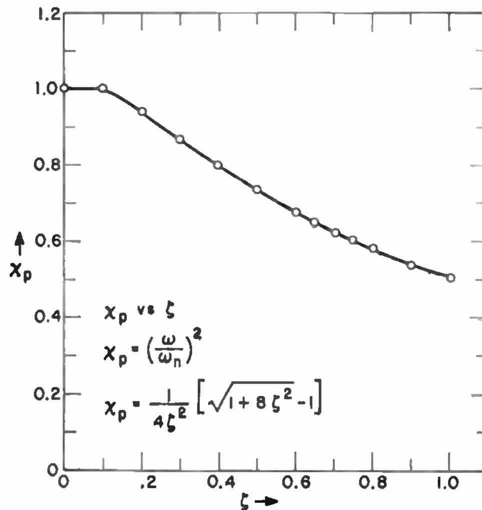


Fig. 7—χ_p versus ζ.

by using the following relations[6]:

$$\frac{2B_L}{2B_{L_0}} = \frac{1}{3}\left(1 + \frac{2}{\mu}\right), \tag{11}$$

$$\mu = \alpha/\alpha_0 . \tag{12}$$

where $2B_{L_0}$ is the loop noise bandwidth at threshold. If either α_0 or $2B_{L_0}$ is known, the other can be determined using Equations (11) and (12). It is assumed that $2B_L$ is known for a strong signal. It is virtually impossible, however, to determine $2B_{L_0}$ experimentally, because the loop maintains lock for only a very small percentage of the time. α_0 can be determined by eliminating $2B_{L_0}$ as follows:

$$\frac{2B_{L_1}}{2B_{L_0}} = \frac{1}{3}\left(1 + \frac{2}{\mu_1}\right), \tag{13}$$

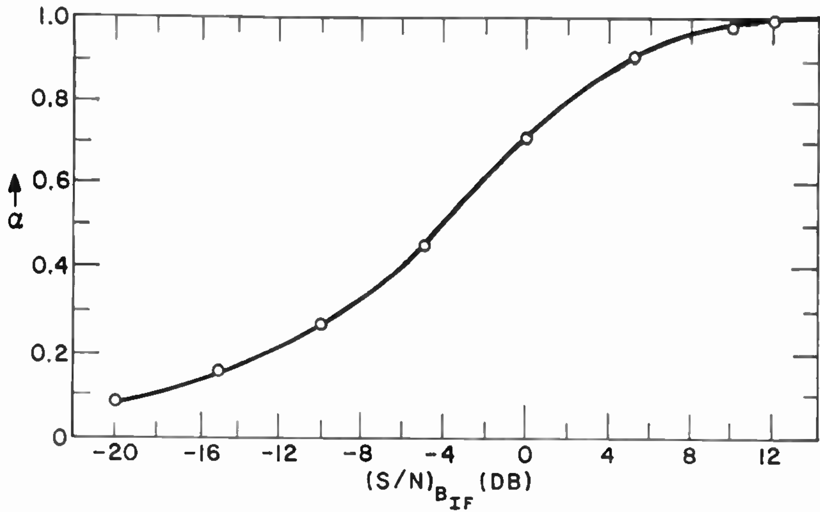


Fig. 8—The suppression factor (α) for an ideal limiter.

$$\frac{2B_{L_2}}{2B_{L_0}} = \frac{1}{3}\left(1 + \frac{2}{\mu_2}\right). \tag{14}$$

Dividing Equation (14) into Equation (13) yields

$$\frac{2B_{L_1}}{2B_{L_2}} = \frac{\alpha_0 + 2\alpha_2}{\alpha_0 + 2\alpha_1} \tag{15}$$

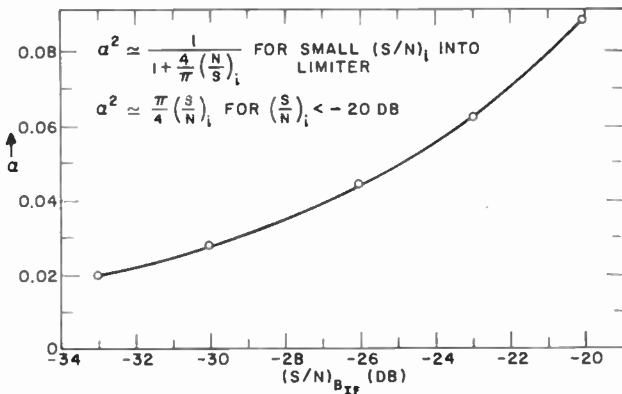


Fig. 9—Expanded region of limiter suppression factor α.

Substitution of values from Figures 4 and 5 results in $\alpha_0 = 0.0587$ for the particular loop under study. A check of this value by a process of iteration yielded $\alpha_0 = 0.056$, or a descrepancy between the two values of only 4.6%, indicating the reliability of the experimental determination.

The process of interation is as follows:

(a) Assume a value of $2B_{L_0}$. As a start, $2B_L$ for a strong signal, which is assumed known (or can be found by measurement), may be used.

(b) Find $(S/N)_{B_{if}} = \left(\frac{S}{N}\right)_{2B_{L_0}} \frac{2B_{L_0}}{B_{if}}$.

Note that $\left(\frac{S}{N}\right)_{2B_{L_0}} = 0$ db is the absolute threshold point, by definition.[6]

(c) Find α_0 using either Figure 8 or 9.

(d) Find a new value of $2B_{L_0}$ using

$$\frac{2B_L}{2B_{L_0}} = \frac{1}{3}\left(1 + \frac{2}{\alpha_0/\alpha}\right).$$

Here $\alpha = 1$ at strong signal, and $2B_L$ for a strong signal is also known.

(e) Reiterate the previous steps until there is a convergence on values for $2B_{L_0}$ and α_0.

When the value $\alpha_0 = 0.0587$ is inserted in Equations (11) and (12), they yield the bandwidth at threshold for the specific loop under consideration. Thus, since $2B_L = 1500$ Hz for this loop,

$$\mu = \frac{\alpha_0}{\alpha} = \frac{0.0587}{1},$$

$$\frac{1500}{2B_{L_0}} = \frac{1}{3}\left(1 + \frac{2}{0.0587}\right),$$

and

$$2B_{L_0} = 130 \text{ Hz}.$$

SIGNAL-TO-NOISE RATIO IN THE LOOP BANDWIDTH VERSUS SIGNAL-TO-NOISE RATIO IN THE I-F BANDWIDTH

From the experimentally established values of α_0, $2B_{L_0}$, and $(S/N)_{B_{if}}$ at absolute threshold (corresponding to α_0 in Figure 9), the $(S/N)_{2B_L}$ can be found using Figure 10. This figure yields, for any α

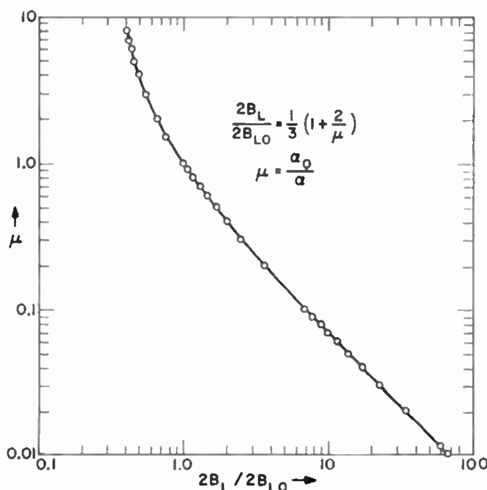


Fig. 10—$2B_L/2B_{L_0}$ versus μ.

other than α_0, a corresponding $2B_L$. Knowing $2B_L$ and B_{if}, which is 30 kHz in this case, the $(S/N)_{2B_L}$ can be found;

$$(S/N)_{2B_L} = (S/N)_{B_{if}} \bigg|_{\text{at } \alpha_0} \times \frac{B_{if}}{2B_L}. \tag{16}$$

The result of this calculation is shown in Figures 11 and 12. Figure 11 gives the variation of $(S/N)_{2B_L}$ as a function of the double-sided loop noise bandwidth, $2B_L$, for the loop under consideration. The range of signal-to-noise ratio covered by the figure is more than adequate for the purpose of the mean-time-to-unlock measurements to be discussed in the following section.

Figure 12 indicates that there is no one-to-one correspondence between $\Delta(S/N)_{B_{if}}$ and $\Delta(S/N)_{2B_L}$. The relationship is essentially nonlinear, but it can be subdivided into the following four linear regions. Here, all quantities are in db's, and pertain to the loop under consideration.

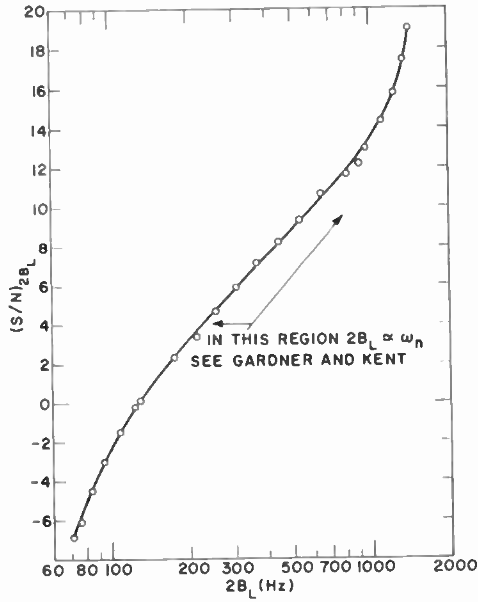


Fig. 11—$(S/N)_{2B_L}$ versus $2B_L$ for the loop.

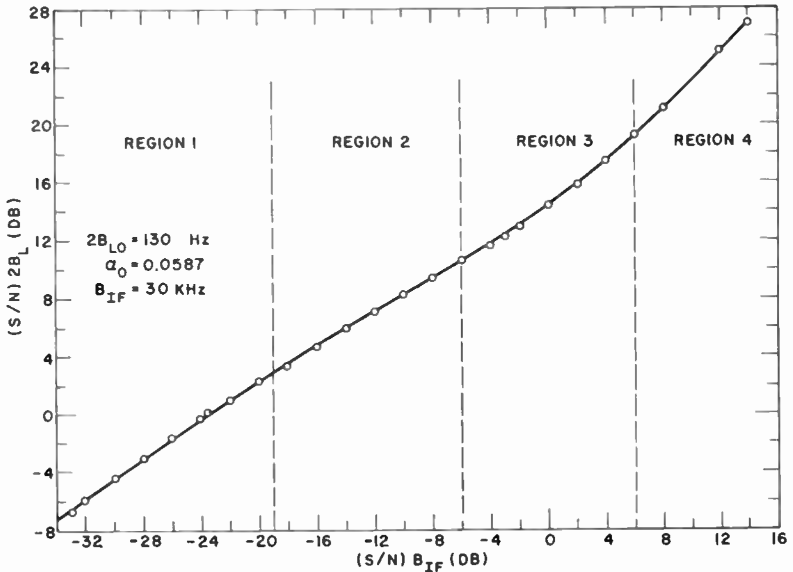


Fig. 12—$(S/N)_{2B_L}$ versus $(S/N)_{B_{if}}$.

Region 1: $-34 < (S/N)_{B_{if}} < -19$ or $-7 < (S/N)_{2B_L} < 3$. In this region, the equivalent linear relationship is

$$(S/N)_{2B_L} \simeq 0.65\ (S/N)_{B_{if}} + 15\,. \tag{17}$$

Region 2: $-19 < (S/N)_{B_{if}} < -6$ or $3 < (S/N)_{2B_L} < 11$. The relationship is adequately described by

$$(S/N)_{2B_L} = 0.588\ (S/N)_{B_{if}} + 14.23\,. \tag{18}$$

Region 3: $-6 < (S/N)_{B_{if}} < 6$ or $11 < S/N)_{2B_L} < 19$. Here, use can be made of the approximate relation

$$(S/N)_{2B_L} = 0.75\ (S/N)_{B_{if}} + 15\,. \tag{19}$$

Table I

Region	$\Delta\ (S/N)_{2B_L}$ Corresponding to $\Delta\ (S/N)_{B_{if}} = 1$ db
1	0.65
2	0.5888
3	0.75
4	1.0

Region 4: $(S/N)_{B_{if}} > 6$ or $(S/N)_{2B_L} > 19$. The straight line relationship is

$$(S/N)_{2B_L} = (S/N)_{B_{if}} + 13\,. \tag{20}$$

Thus an incremental change of $(S/N)_{B_{if}}$ by 1 db produces different incremental changes in $(S/N)_{2B_L}$ depending on the region of operation. They are enumerated in Table I. One-to-one correspondence is obtained in region 4 where the signal is relatively strong, which is in agreement with theoretical investigations.

MEASUREMENT TECHNIQUE FOR THE UNLOCK CHARACTERISTICS

The experimental setup is shown in Figure 13. The phase-lock test receiver No. 2 is driven by a strong signal and is in lock at all times. Phase lock receiver No. 1 is driven through an attenuator so that the signal-to-noise ratio can be controlled to be within the region where the technique yields best results. The outputs of the loop voltage-controlled oscillators of both receivers are summed and then correlated by passing them through a diode. If we let the output of the continuously locked oscillator be $A \cos \omega_0 t$ and the output of the intermittently un-

locking oscillator be $B \cos \omega t$, then the output of the diode can be easily shown to be of the form

$$K_1 \cos \omega_o t + K_2 \cos \omega t + K_3 \cos 2\omega_o t + K_4 \cos 2\omega t$$
$$+ K_5 \cos (\omega_o + \omega) t + K_6 \cos (\omega_o - \omega) t + K_7 .$$

The low-pass filter provides filtering of the higher-frequency components and transmission of the frequency component of interest, $(\omega_o - \omega)$, in addition to the d-c component K_7. Thus, when both

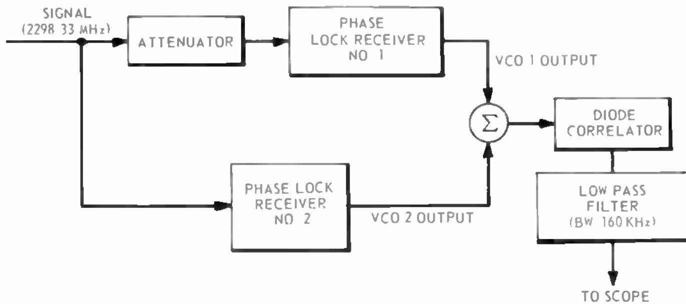


Fig. 13—Block diagram representation of the experimental technique used to measure unlock characteristics.

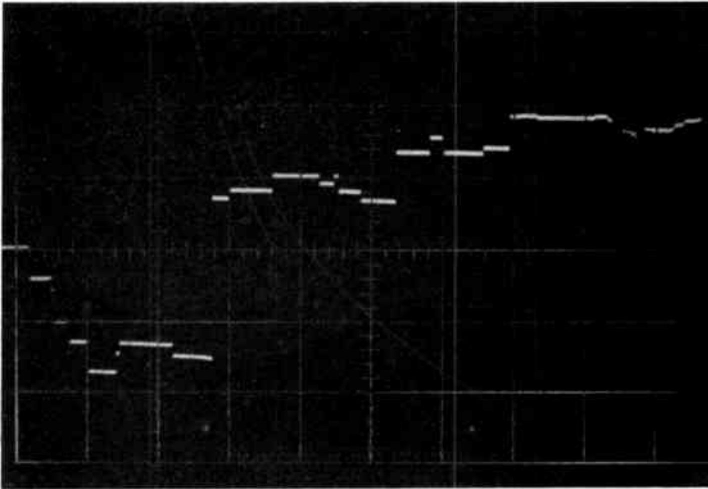
voltage-controlled oscillators are in lock, the output of the low-pass filter is d-c, i.e., $K_6 + K_7$. When they are out of lock, the difference frequency $(\omega_o - \omega)$ is observable on the scope.

If the lightly driven VCO skips a few cycles due to noise bursts but locks on again, the difference $(\omega_o - \omega)$ during the "skip time" is displayed. The incremental phase difference between the two oscillators corresponding to the two consecutive conditions of lock is displayed as a difference in d-c levels. Similarly, if due to noise the lightly driven VCO skips a fraction of a cycle, the jump in phase, $\theta_o - \theta_1$, is displayed on the scope by a jump in d-c level proportional to $\cos (\theta_o - \theta_1)$.

Both voltage-controlled oscillators operate at 23.635 MHz, while the input signal frequency is 2298.3 MHz. The frequency multiplication and mixing are performed within the test receiver.

Thus, if x cycles are skipped at the r-f level, the VCO skips $x/96$ cycles (see Figure 1). If one cycle is skipped at the r-f level, the phase of the lightly driven VCO jumps by $\pm 360°/96 = \pm 3.75°$, and thus the d-c level changes instantaneously by an amount proportional to the

cosine of that angle. Because there is only one stable point in each cycle, the minimum possible phase jump is 3.75°, and all other phase jumps are integral multiples of 3.75°. These phase jumps are recorded on a photograph such as that shown in Figure 14. This technique was used to measure at least 100 different periods of continuous lock for each signal-to-noise ratio. The mean time to unlock was then obtained on the basis of these one hundred samples.



Horizontal scale: 2 sec/cm — $(S/N)_{2B_L} = 6.6$ db

Vertical scale: 0.1 volt/cm — $(S/N)_{B_{if}} = -13$ db

$2B_L \simeq 350$ Hz

Fig. 14—Unlock behavior for second-order loop for single carrier.

An interfering carrier displaced in frequency by a Δf Hz was added to the original carrier at different relative amplitudes and the same technique was used for finding the mean time to unlock. The frequency Δf was measured in the loop at the input to the VCO.

Results

The normalized mean time to unlock for a single carrier at different signal-to-noise ratios in the loop bandwidth, $(S/N)_{2B_L}$, is plotted in Figure 15. To normalize, the mean time to unlock in seconds was multi-

plied by ω_n. For the range over which the measurements were taken Reference (5) indicates that

$$\omega_n \simeq 2B_L . \tag{21}$$

This relation was used to compute the normalized mean time to unlock.

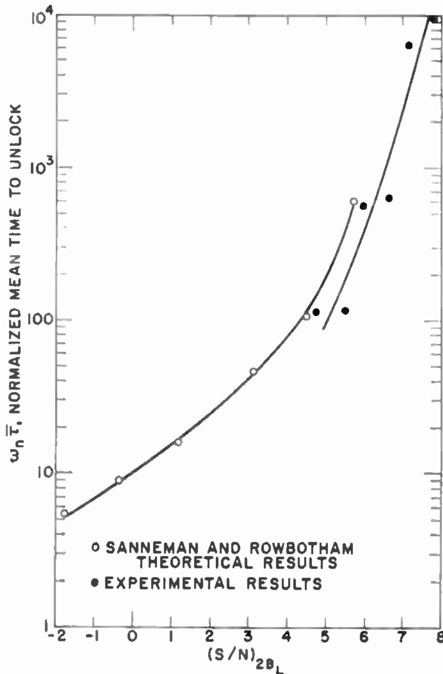


Fig. 15—Normalized mean time to unlock for single carrier.

Figure 15 indicates that the experimental technique yields results that are compatible with theoretical results as obtained by Sanneman and Rowbotham.[1] In Sanneman's theory, the system is assumed perfectly linear; $\zeta = 0.707$ and $2B_L = 1.06\omega_n$. Over the experimental range, ζ varied between 0.4 and 0.6. This variation does not affect the results appreciably, since in that range[5] $\omega_n \simeq 2B_L$. Sanneman's results can therefore be compared to the experimental results in this range. The linearization does not hold for $(S/N)_{2B_L}$ near 0 db and any experimental results will not, strictly speaking, be compatible with Sanneman's theory.

It should be mentioned here that the theoretical curve in Figure 15 was obtained from Sanneman's results by using the approximation, for the noise phase jitter σ_p,

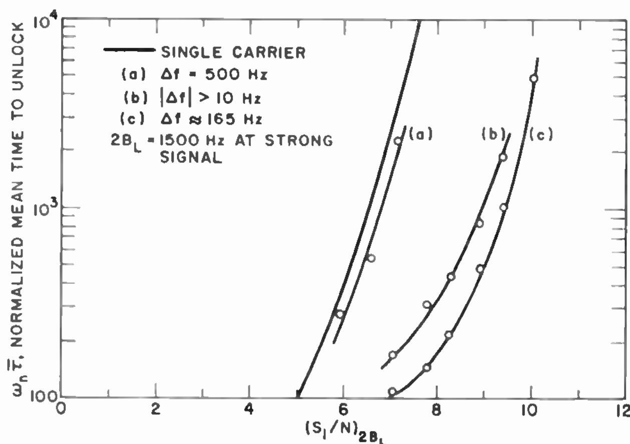


Fig. 16—Normalized mean time to unlock for interferring carriers 3 db below the main carrier.

$$\sigma_p^2 = \frac{1}{(S/N)_{2B_L}} \tag{22}$$

which is true[6] for $(S/N) < 10$.

Figure 16 is a plot of the NMTU against the $(S_1/N)_{2B_L}$, where S_1 is the power of the tracked carrier. The three different graphs are for an interfering carrier whose power is S_2 such that $S_1 - S_2 = 3$ db, with different frequency separations between the carriers. For comparison, the NMTU for a single carrier is also shown. Figure 17 shows the

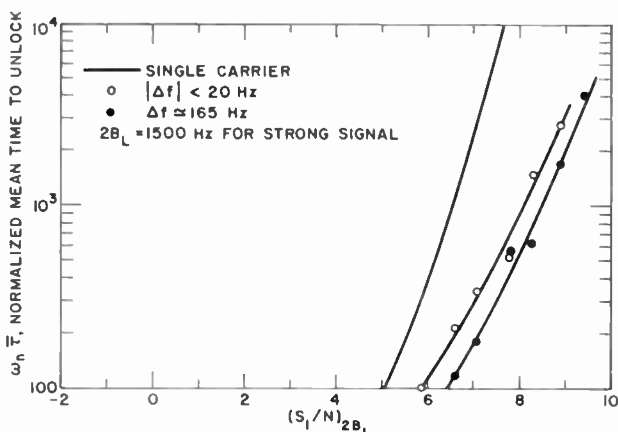


Fig. 17—Normalized mean time to unlock with interferring carriers 5 db below the main carrier.

NMTU versus $(S/N)_{2B_L}$ for a single carrier and for an interfering carrier such that $S_1 - S_2 = 5$ db, with different frequency separations. In Figure 18, $S_1 - S_2 = 10$ db.

Because the loop noise bandwidth is variable with signal level, there is a need to show the NMTU versus a signal-to-noise ratio in a constant bandwidth. This is provided in Figure 19, where the bandwidth is taken as B_{if}.

CONCLUSIONS

The experimental results for a single carrier perturbed by white gaussian noise are within the order of magnitude theoretically predicted by Sanneman.[1] In particular, the experimental loop behavior within the region $10^2 < \text{NMTU} < 10^4$ coincides with the theoretical results to within the experimental accuracy. The experimental technique is not suited for measurements of $\text{NMTU} < 10^2$ and therefore comparison in that range is not possible.

On the other hand it is not possible to extract meaningful information from the photographic technique for $\text{NMTU} > 10^4$. The limitations imposed by the photographic technique of recording does not impair the basic correlation technique used to detect unlock. Thus, counters and other recording devices may be used in the NMTU regions where either may be applicable.

It is clear from Figure 16 that an interfering carrier with a frequency separation of 500 Hz and an amplitude separation as large as 3 db below the tracked carrier has negligible effect on the unlock characteristics of the loop tracking a carrier perturbed by noise. An interfering carrier with a frequency separation of 165 Hz may require an increase of S_1 by 2 to 3 db in order to produce the same NMTU. For the same $(S/N)_{2B_L}$, the degradation of NMTU with an interfering carrier may be anywhere between one or two orders of magnitude. Curve (b) of Figure 16 was taken for $-10 \text{ Hz} < \Delta f < 10 \text{ Hz}$. Since Δf can take any value in that range, both carriers will have the same frequency part of the time and hence the tracked carrier is enforced. This causes the noise bursts to be less effective in unlocking the loop. The NMTU for $|\Delta f| > 10$ Hz is then larger than that for $\Delta f \approx 165$ Hz, as shown in Figure 16.

The same type of behavior is observable in Figure 17 for interfering carriers whose level is 5 db below the tracked carrier. The frequency difference between the carriers is 20 Hz and 165 Hz. The order of magnitude of the performance degradation is less, however, since the interfering carrier is smaller in magnitude. For the same $(S/N)_{2B_L}$, the degradation may be anywhere between 0.5 and 1.5 orders of mag-

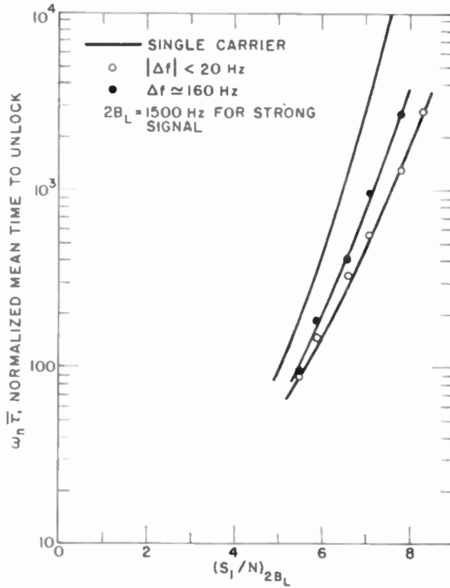


Fig. 18—Normalized mean time to unlock with interferring carriers 10 db below the main carrier.

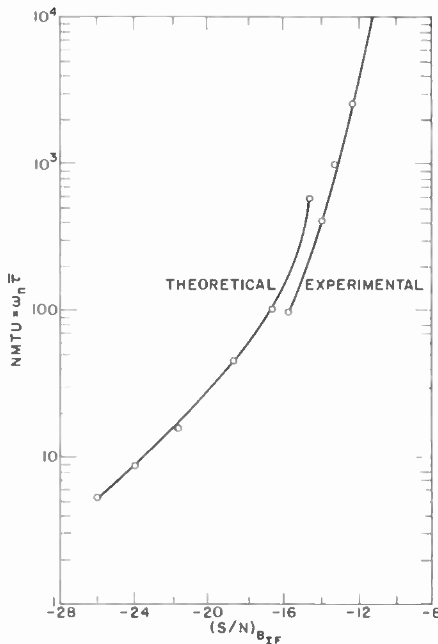


Fig. 19—Normalized mean time to unlock for a single carrier versus S/N in the i-f bandwidth.

nitude. Equivalently, to obtain the same NMTU as in Figure 15, the level of the tracked carrier may have to be increased by 1.5 to 2.5 db.

In Figure 18 the maximum degradation is one order of magnitude; it remains less than that for the rest of the experimental data range. In this figure the results for -20 Hz $< \Delta f <$ 20 Hz do not display improvement over the 160 Hz difference-frequency curve. The reason for this remains unresolved.

One important facet of the loop behavior emphasized by the present investigation is the nonlinear relationship between $(S/N)_{2B_L}$ and $(S/N)_{B_{if}}$. The relationship is linear only for strong signals. It is important also to exercise extreme care when determining the signal-to-noise ratio at threshold in the loop bandwidth because of the fact that the loop bandwidth changes with signal level.

In the instance where the mean time to unlock is an important loop parameter, the $(S/N)_{2B_L}$ corresponding to that parameter can be determined directly by the use of Figure 15 for no interfering carrier and by interpolation of the curves of Figures 16, 17, and 18. For the designer, however, Figure 19 provides the NMTU for the $(S/N)_{B_{if}}$, where $B_{if} = 30$ kHz. This can be simply translated into a (S/N) in a different i-f bandwidth when desired.

Figures 16, 17, and 18 indicate that for a certain signal level and for a certain NMTU, there is a frequency of interfering carrier at which the $(S_1/N)_{2B_L}$ is maximum. Further study in this respect seems justifiable.

ACKNOWLEDGEMENTS

The author expresses his appreciation for the Laboratory assistance given by H. Rickert and J. Holmquest. The author also acknowledges the assistance of H. R. Mathwich for his appraisal of the technique and for his interepretation of some aspects of the results.

LANDAU DAMPING IN ELECTRON BEAMS

By

MARTIN CAULTON

RCA Laboratories
Princeton, N. J.

Summary—It is shown that the exponential damping rate predicted by Landau for a stationary plasma with a Maxwellian distribution of velocities is equivalent to the spatial damping rate calculated for a moving electron beam with a cut-off Maxwellian distribution of velocities. This is done (1) by extracting the spatial damping rates and converting them to temporal equivalents, and (2) by using a Galilean transformation and a Doppler frequency shift to show that the damping-determining factors of the moving electron beam are equivalent to the damping-determining factors for a stationary plasma, i.e., the ratio of plasma wave number to Debye wave number. The damping rates of over 50 experiments reported earlier involving the decay of space-charge waves in electron beams are then found to agree with the predictions of the electron beam theories, i.e., Landau damping. It is suggested that this constitutes a quantitative verification of Landau damping in electron beams.

INTRODUCTION

IT HAS been shown experimentally[1,2] that space-charge waves, excited in electron beams with large velocity spreads, decay with distance along the beam. Berghammer[3] and Bloom and Vural[4,5] have made theoretical studies of an electron beam with a cut-off Maxwellian distribution of axial velocities. Their theory (referred to here as the BBV theory) predicts the damping of one of the Hahn–Ramo space-charge waves. The resultant r-f power decay with distance is similar to the experimental results mentioned. It is claimed that this decay is that predicted by Landau,[6] who studied the decay of waves

[1] M. Caulton, B. Hershenov, and F. Paschke, "Experimental Evidence of Landau Damping in Electron Beams," *Jour. Appl. Phys.*, Vol. 33, p. 800, March 1962.

[2] T. G. Mihran, "RF Current Behavior in Electron Beams with DC Velocity Spread," *Jour. Appl. Phys.*, Vol. 33, p. 1582, April 1962.

[3] J. Berghammer, "Landau Damping of Space-Charge Waves," *Jour. Appl. Phys.*, Vol. 33, p. 1499, April 1962.

[4] S. Bloom and B. Vural, "Noise on a Drifting Maxwellian Beam," *Jour. Appl. Phys.*, Vol. 34, p. 356, Feb. 1963.

[5] S. Bloom and B. Vural, "Space-Charge-Wave Decay Along a Signal-Current-Excited Multivelocity Beam," *Jour. Appl. Phys.*, Vol. 34, p. 2007, July 1963.

[6] L. P. Landau, "On the Vibrations of the Electronic Plasma," *Jour. Phys.* U.S.S.R., Vol. 10, p. 25, 1946.

with time in a stationary plasma with a Maxwellian distribution of velocities. The claim that the r-f power decay in a current-modulated electron beam is Landau damping has been questioned.[7,8] A recent paper by the author[9] describes an experimental quantitative study of the damping of waves in electron beams. The values of the damping rates are in agreement with the predictions of the BBV analysis. It is the purpose of this paper to show that the predictions of this theory lead to a damping rate of equivalent to Landau's. This is done by showing that when the BBV spatial damping in a moving electron beam is transformed into a temporal damping in a stationary plasma, the result is the same as that predicted by Landau. Thus the damping rates measured in electron beams can be considered a quantitative verification of Landau damping in electron beams.

LANDAU DAMPING

The BBV and Landau theories are compared in the framework of the temporal damping of a stationary plasma as developed by Jackson.[10] The linearized, collisionless, one-dimensional Boltzmann equation yields the dispersion relationship:

$$\frac{k^2}{\omega_p{}^2} = \int_{-\infty}^{\infty} \frac{\dfrac{1}{n_o}\dfrac{\partial f_o}{\partial v}}{v - \omega k}\, dv\,, \tag{1}$$

where v is the velocity,

$f_o(v)$ is the unperturbed velocity distribution,

n_o is the average number of particles per unit volume,

ω is the frequency of oscillation,

k is the wave number or propagation constant,

ω_p is the plasma frequency,

---

[7] T. G. Mihran and S. P. Yu, "Nonlinear RF Behavior of Electron Beams with Velocity Distribution .II—Application to Rectangular Velocity Distribution," *Jour. Appl. Phys.*, Vol. 34, p. 2976, Oct. 1963.

[8] R. L. Gunshor, "Landau Damping in Plasmas with Finite Transverse Dimensions," *Jour. Appl. Phys.*, Vol. 37, p. 3630, Aug. 1966; Vol. 37, p. 1904, 1966.

[9] M. Caulton, "Damping of Waves in Electron Beams," *Jour. Appl. Phys.*, Vol. 38, p. 1839, March 1967.

[10] J. D. Jackson, "Longitudinal Plasma Oscillations," *Jour. Nucl. Energy*, Part C, Plasma Physics, Vol. 1, p. 171, 1960.

e is the charge of an electron,

m is the mass of an electron,

ϵ_o is the permittivity of free space.

Equation (1) in principle can be used to find the frequency ω as a function of k for any given velocity distribution $f_o(v)$. If the problem is specified in terms of an initial-value problem as was done by Landau, we find that ω is made up of a real part ω_r and an imaginary part ω_i, for real values of k;

$$\omega = \omega_r + j\,\omega_i\,. \tag{2}$$

For small values of k the denominator in the principal value of the integral of Equation (1) can be expanded to show that the real part of ω is approximately given by

$$\omega_r^2 \cong \omega_p^2 + 3\,k^2 < v^2 >\,, \tag{3}$$

where $< v^2 >$ is the mean squared velocity of the assumed symmetric distribution $f_o(v)$.

Jackson also derives a general expression for the imaginary part of ω (ω_i) that causes damping that is valid for $\omega_i << \omega_r$:

$$\omega_i \cong \frac{\pi}{2}\,\omega_r \left(\frac{\omega_p}{k}\right)^2 \frac{1}{n_o} \left(\frac{\partial f_o}{\partial v}\right)_{v = \omega_r/K} \cdot \left(1 - \frac{k}{\omega_r}\,\frac{d\omega_r}{dk}\right) \tag{4}$$

We shall evaluate Equation (4) both for a stationary plasma and for a moving beam.

The Stationary Plasma

A one-dimensional Maxwellian velocity distribution of a stationary plasma in thermal equilibrium is

$$f_o(v) = n_o \frac{1}{\sqrt{2\,\pi}\,C} \exp\left\{-\frac{v^2}{2C^2}\right\}, \tag{5}$$

where $C = \sqrt{KT/m}$

K is Boltzmann's constant,

T is the equivalent temperature of the distribution.

For a full symmetric Maxwellian distribution such as Equation (5) the mean squared velocity $< v^2 >$ is identically C^2 or KT/m. Then

$$\omega_r^2 \cong \omega_p^2 + 3\, k^2 C^2 \,. \tag{6}$$

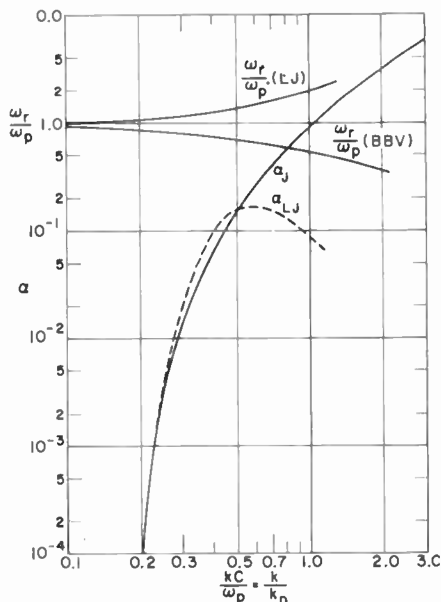


Fig. 1—Exponential damping rates α (from Reference (10)) and ω_r/ω_p as function of k/k_D.

Using Equation (6) and the velocity distribution of Equation (5), Jackson derives from Equation (4) an expression for the Landau damping constant, called here α_{LJ};

$$\alpha_{LJ} = \frac{\omega_i}{\omega_p} = -\sqrt{\frac{\pi}{8}}\left(\frac{\omega_p}{KC}\right)^3 \exp\left\{\frac{-\omega_p^2}{2\,k^2C^2} - \frac{3}{2}\right\}. \tag{7}$$

Figure 1 shows α_{LJ} as a function of kC/ω_p; this is reproduced from Jackson's[10] Figure 3. Also shown in Figure 1 are the results of a more exact expression derived by Jackson, called here α_J, which is a better approximation when ω_i and kC/ω_p are large. The real part of ω_r/ω_p from Equation (6) is also plotted. We note that the approximate term α_{LJ} coincides with the more exact α_J only for small kC/ω_p, since α_{LJ} is limited by the validity restrictions of Equations (4) and (6). The

abscissa of Figure 1, kC/ω_p, can also be expressed as k/k_D, where k_D is the Debye wave number ω_p/C.

In the preceding, ω_p is the plasma frequency for an infinite plasma. For a plasma with finite dimensions, such as in an electron beam, the plasma frequency ω_p is reduced to ω_q. The reduction factor ω_q/ω_p is a function of k and drift tube and beam diameters. However, the form of the Equation (7) would be unchanged for the finite plasma, and Figure 1 should still give the same results in terms of ω_q. Gunshor[8] derives Equation (7) in terms of ω_p, taking into account ω_q/ω_p, and shows the large increase in the damping constant α_{LJ} that results. This increase appears only because the damping is compared with the infinite-beam plasma frequency ω_p.

THE ELECTRON BEAM OR MOVING PLASMA

Let us now examine the moving plasma. The papers by Berghammer and by Bloom and Vural in *principle* solve the problem of finding ω_r and the damping constant ω_i. Figure 2 is a schematic representation of the predictions of the BBV theory. In Figure 2(a), a univelocity beam that is velocity modulated is indicated as giving rise to equal-amplitude fast and slow space-charge waves. The velocity modulation on the univelocity beam sets up plasma oscillations at the plasma frequency ω_p about the electrons' equilibrium position. The Galilean transformation from the electron equilibrium position to the laboratory system Doppler shifts the plasma frequency ω_p or ω_r to the modulating frequency ω_m, and transforms the waves going forward and backward in the beam coordinate system to the slow and fast waves interfering with one another spatially in the laboratory system.[11] Figure 2(b) and (c) indicates the prediction of BBV for a cut-off Maxwellian distribution. The sharp low-energy cutoff would still give rise to a discrete slow space-charge wave. The exponential high-energy tail, on the other hand, gives rise only to a collection of waves of a continuous frequency spectrum. At the modulating cavity, the integral over this continuous spectrum would give rise to a so-called wave packet. This packet is supported by the electrons in one velocity class only. There is no collective behavior and this packet decays with distance from the modulating cavity via Landau damping. According to this model, a slow wave is left, representing the remaining coherent r-f power in the beam. Figure 2(c) indicates the prediction of the decrease of the amplitude of the remaining wave with increasing velocity spread. It is desired to compare the Landau damp-

[11] C. L. Hemenway, R. W. Henry, and M. Caulton, *Physical Electronics*, p. 318, John Wiley & Sons, New York, 1962.

and Bloom[12] derive a relation for this change in terms of the invarient cathode temperature T and v_o. We express this as

$$\frac{KT_e}{m} v_o^2 = \left(\frac{KT}{m}\right)^2,$$

or (10)

$$Cv_o = \frac{KT}{m}.$$

In the theoretical results of BBV, no analytical value is given for ω_i. However, a solution for ω_i was derived and numerical values are given by BBV from which ω_i can be calculated. Berghammer,[3] for instance, shows damping curves (r-f current squared versus distance for a beam with an initially pure velocity modulation) in terms of a quantity ϵ/λ. Bloom and Vural[5] show similar curves (r-f current versus distance from an initially pure current modulation plane) also in terms of the parameter ϵ/λ. The parameters ϵ and λ are given as

$$\epsilon = \frac{KT}{2\,eV_o}, \qquad (11)$$

$$\lambda = \frac{\omega_{ps}}{\omega_m}, \qquad (12)$$

where ω_m is the modulating or signal frequency in the laboratory reference frame and the plasma frequency ω_{ps} is defined as

$$\omega_{ps}^2 = \frac{ei_o}{m\epsilon_o u_s}. \qquad (13)$$

We can now show that ϵ/λ, which determines the damping in the BBV theory, is nearly equal to k/k_D of the stationary plasma description. The term k is an invarient in the Galilean transformation from the laboratory system to the beam coordinate system.[11] The frequencies of the two systems are Doppler shifted from one another and are related by

$$(\omega_m - ku_s)^2 = \omega_r^2. \qquad (14)$$

---

[12] J. Berghammer and S. Bloom, "On the Nonconservation of Noise Parameters in Multivelocity Beams," *Jour. Appl. Phys.*, Vol. 31, p. 454, March 1960.

Equation (14) gives the wave number for the well-known fast waves (k_f) and slow waves (k_s) of electron beams;

$$k_f = \frac{\omega_m - \omega_r}{u_s}, \tag{15}$$

$$k_s = \frac{\omega_m + \omega_r}{u_s}. \tag{16}$$

For systems where

$$\omega_p << \omega_m ,$$

$$\omega_r << \omega_m ,$$

the propagation constant k is mainly determined by ω_m and u_s imposed in the laboratory system.

Completing the re-expression of ϵ/λ, we have, from Equation (10), (11), and (13) and using the approximation* $v_o \approx u_s$,

$$\frac{C}{v_o} \cong \frac{KT}{2\,eV_o} \equiv \epsilon . \tag{17}$$

Approximating† $\omega_p \approx \omega_{ps}$ for small ϵ, and using Equation (16) for the slow wave, we have

$$\frac{\epsilon}{\lambda} = \frac{kC}{\omega_p} - \frac{\epsilon}{v_o}\frac{\omega_r}{\omega_p} \approx \frac{k}{k_D} \tag{18}$$

for ϵ small compared with v_o, and ω_r of the order of ω_p.

With the equivalence of ϵ/λ and k/k_D established, it might appear that the damping rates exhibited in Figure 1 would apply to the BBV theory. However, the real part of ω, ω_r, is not the same for the full and the cut-off Maxwellian distributions. References (4) and (5) explicitly describe a slow-wave propagation constant s as a function of ϵ/λ. We now derive a new dispersion relation to apply to the cut-off Maxwellian distribution. The slow wave is described by exp $\{-j\epsilon s\theta\}$, where

* More exactly, $v_o/u_s = (1 + \epsilon - 2\epsilon^2 \ldots)$.

† More exactly, $\omega_p^2 = \omega_{ps}^2\,(1 - \epsilon + 3\epsilon^2 \ldots)$.

$$\theta = \frac{\omega_m}{u_s} z \, ,$$

and z is the axial distance along the beam. Then

$$\exp\{-j\epsilon s\theta\} = \exp\left\{-j\frac{\omega_r}{v_o} z\right\}, \tag{19}$$

Thus

$$\omega_r = \frac{\epsilon}{\lambda} s \, \omega_{ps} \frac{v_o}{u_s} . \tag{20}$$

As before we let $\omega_{ps} \approx \omega_p$, and the average velocity $v_o \approx u_s$, so

$$\omega_r \cong \frac{\epsilon}{\lambda} s \, \omega_p . \tag{21}$$

Equation (21) is the dispersion equation that applies to the cut-off Maxwellian distribution of the moving beam and is to be used to derive ω_i for the moving beam. From Reference (4), Equation (27), s is implicitly a function of ϵ/λ given by

$$\left(\frac{\epsilon}{\lambda}\right)^2 = e^s E_i(-s) + \frac{1}{s}, \tag{22}$$

where $E_i(-s)$ is the exponential integral

$$E_i(-s) = -\int_s^\infty \frac{1}{t} e^{-t} \, dt .$$

Let us now use the dispersion relation Equation (21) for ω_r as derived by BBV, along with the Galilean transformation discussed earlier, in the general expression for ω_i derived by Jackson, Equation (4). Evaluating the terms of Equation (4)

$$\frac{d\omega_r}{dk} = \frac{d\dfrac{\epsilon}{\lambda}}{dk} s \, \omega_p + \frac{\epsilon}{\lambda} \omega_p \frac{ds}{d\dfrac{\epsilon}{\lambda}} \frac{d\dfrac{\epsilon}{\lambda}}{dk} + \frac{\epsilon}{\lambda} s \frac{d\omega_p}{dk} . \tag{23}$$

The term $d\omega_p/dk$ is zero for an infinite beam. For a finite confined beam, the equivalent $d\omega_q/dk$ would be negligible for large values of $\omega_m b/u_s$, where b is the beam radius. Using Equations (18) and (22), we have

$$1 - \frac{k}{\omega_r}\frac{d\omega_r}{dk} = \frac{2}{s}\frac{\left(\frac{\epsilon}{\lambda}s\right)^2}{1-\left(\frac{\epsilon}{\lambda}s\right)^2}. \tag{24}$$

We obtain a modified damping constant we call α_M, which is the moving-beam BBV Landau damping constant applied to a stationary plasma. This is found by using Equation (24) and the f_o for a stationary Maxwellian distribution of Equation (4). The use of the stationary distribution function of Equation (5) is plausible. It is equivalent to examining the cut-off Maxwellian distribution of Equation (8) from a system moving at velocity u_s along with the beam and seeing only the faster electrons with positive velocities.* We are then only describing the damping constant for the wave associated with the faster electrons. The transformation invarients and the Doppler-shifted frequency are as described earlier. Then

$$\alpha_M = -\sqrt{\frac{\pi}{8}}\left(\frac{\omega_p}{kC}\right)^3 \exp\left\{-\left(\frac{\epsilon}{\lambda}s\right)^2\frac{\omega_p^2}{2k^2C}\right\}\cdot\frac{2}{s}\frac{\left(\frac{\epsilon}{\lambda}s\right)^4}{1-\left(\frac{\epsilon}{\lambda}s\right)^2}. \tag{25}$$

This can be expressed in terms of the Landau damping rate α_{LJ} described by Jackson;

$$\alpha_M = \alpha_{LJ}\frac{2}{s}\frac{\left(\frac{\epsilon}{\lambda}s\right)^4}{1-\left(\frac{\epsilon}{\lambda}s\right)^2}\exp\left\{-\frac{1}{2}\left[s^2-\left(\frac{\lambda}{\epsilon}\right)^2\right]+\frac{3}{2}\right\} \tag{26}$$

* We are also neglecting the ions that are stationary in the laboratory system. In the beam system they are back streaming with velocities greater than the thermal velocity. They should not interact, although their presence is necessary to neutralize the average space charge.

We use Equation (26) to obtain the equivalent temporal damping rate α_M for a moving beam. In Figure 4, α_M is plotted for ϵ/λ from 0.2 to 1.0.

Comparisons of Theories of Moving and Stationary Plasma

The solid line of Figure 4 is the damping rate we call α_{BV} extracted from Bloom and Vural[5] (Figure 4) as described in Appendix I. Three

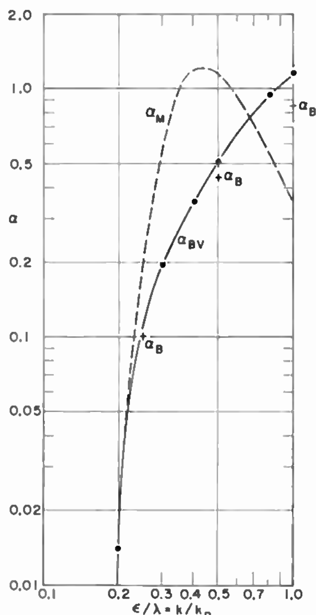


Fig. 4—Exponential damping rate α as function of $\epsilon/\lambda = k/k_D$. α_M is the stationary plasma theory modified by Reference (4) dispersion term, α_{BV} is the electron beam theory extracted from Reference (5), and α_B is the electron beam theory extracted from Reference (3).

points α_B found from Berghammer's Figures 3, 4, and 5 are also shown. These differ slightly from Bloom and Vural's determinations found by a different approximation. It is noted that α_M rises above α_{BV} for high ϵ/λ or k/k_D, similar to the manner in which α_{LJ} rises above Jackson's more accurate α_J. More importantly, we see that α_{BV} and α_B of the moving-beam theory approaches α_M of the stationary plasma theory at small values of ϵ/λ. It is only at these small values of ϵ/λ or k/k_D that α_M of Figure 4 or α_{LJ} of Figure 1 would be valid. The agreement of α_M and α_{BV} at $\epsilon/\lambda = 0.2$ is considered very significant, since the correction factor (Equation (26)) to α_{LJ} of Figure 1 to obtain α_M of Figure 4 is

almost 2 orders of magnitude. This correction brings α_{LJ} into agreement with α_M. Of course, the damping rate for the electron beam applies to only one of the Hahn–Ramo waves.

The real part of ω, ω_r, can be solved for as a function of ϵ/λ using Equations (21) and (22). In Figure 1, $(\omega_r/\omega_p)_{BBV}$ is plotted and the decrease of ω_r with increasing ϵ/λ is noted. It is very interesting to observe that the full Maxwellian distribution, as well as the rectangular case, predicts an ω_r that increases with k/k_D (Equation 3), whereas the half-Maxwellian distribution used for beams predicts a decreasing ω_r. This latter decrease was also found by the analysis of Watkins[13] and by an exact computer analysis of Morrison.[14] It is noted that these studies of the cut-off Maxwellian[12-14] consider the lower edge of the distribution to be of constant velocity, u_s, while the rectangular and full Maxwellian studies consider the average velocity, v_o, to be constant. This consideration does not alter the results of the previous calculations, i.e., a rectangular distribution predicts[15] an increase in ω_r and a cut-off Maxwellian a decrease in ω_r with increasing velocity spread, regardless of which velocity (v_o or u_s) is kept constant. Although the different behavior was not studied in detail, it is felt to be due to the difference (assymmetry) in velocity distribution functions. The increase in ω_r and corresponding decrease in interference-pattern wavelength is more likely to be observed experimentally, because a cut-off Maxwellian distribution would often be masked by velocity spreads that cause an increase in plasma frequency.

Comparison with Experiment

It is concluded that the electron-beam theory of BBV predicts the same exponential damping rate for the small ϵ/λ or k/k_D limit as does the approximate solution developed for the Landau theory of a stationary plasma. On the other hand α_M holds for large ϵ/λ and is limited only to small ϵ. The BBV results would therefore be expected to more accurately predict the r-f damping behavior in electron beams (that can be represented by an equivalent cut-off Maxwellian distribution), than the approximations* developed from the closed-form solutions of the stationary plasmas.

[13] D. A. Watkins, "The Effect of Velocity Distribution in a Modulated Electron Stream," *Jour. Appl. Phys.*, Vol. 23, p. 568, May 1952.

[14] J. A. Morrison, "Noise Propagation in Drifting Multivelocity Electron Beams," *Jour. Appl. Phys.* (Letter), Vol. 31, p. 2066, Nov. 1960.

[15] G. A. Gray, "Single-Velocity Equivalents for Multivelocity Electron Streams," *Jour. Appl. Phys.*, Vol. 31, p. 370, Feb. 1960.

* The more accurate calculations developed by Jackson for large damping cannot readily be put in a form applicable to an electron beam.

Normal electron beams do not have the cut-off Maxwellian distribution described by Equation (8) and Figure 3(a). Figure 3(b) is a typical measured voltage $F(V)$ for electron beams.[9] It was found in the experiments described in Reference (9) that the electrons going slower than the average (shaded region), rather than the faster electrons, determined the measured damping rates. The BBV theory derived for Figure 3(a), Equation (8), has not yet been rigorously determined to apply to the distribution with a low-energy tail, but certain features can be inferred. The highest-energy electrons in Figure 2, (d) and (e) (region C), have too large a slope of $df(u)/du$ to have any measurable damping. If we do not consider the highest-population region (B), we can also infer that the low-energy tail (region A) (very close experimentally[9] to an equivalent Maxwellian distribution) should give rise to a decaying slow-wave packet. The power remaining in the beam would then be that of the fast wave. A phenomenological proof that the theory of BBV will be appropriate to compute the damping for distributions of the type represented in Figure 2, (d) and (e), with slow-velocity tails is given in Appendix II.

A more complete calculation has been carried out by L. S. Nergaard.[16] He shows that the low-velocity tail produces damping of the slow space-charge wave equivalent to the damping of the fast space-charge wave for a high-velocity tail. He solves the Volterra equation for an arbitrary distribution that can represent the half-Maxwellian distribution of BBV or the distribution functions experimentally measured; the damping results are equivalent.

With the considerations described, it was possible to find for over 50 experiments a value of the constant exponential damping rate α versus the damping determining factor ϵ/λ. Here α was computed from the measured spatial damping rate α' and ϵ/λ was computed from the measured voltage distribution of the slow-energy tail (shaded area of Figure 3(b)). Figure 5 presents the results of the analyses of Reference (9) on a logarithmic plot of α versus ϵ/λ. Various beam experiments are listed by different symbols.

Figure 5 also exhibits the theoretical plots of α versus ϵ/λ (obtained by the author (Figure 4)) from the predictions of Bloom and Vural[15] (α_{BV}) and Berghammer[3] (α_B). These predicted values are in good agreement with the analyzed data. While the individual points have a relatively large spread about the theoretical curves, the bulk of the data follows the variations predicted by the BBV theory. More importantly, the absolute damping values derived cluster about the predicted curves.

---

[16] L. S. Nergaard, RCA Laboratories, Princeton, N. J.; private communication.

In view of the irregularities in the beam profile and the distribution across the beam, the agreement found is remarkably good and within the experimental error. For large ϵ/λ, the α's are not as high as pre-

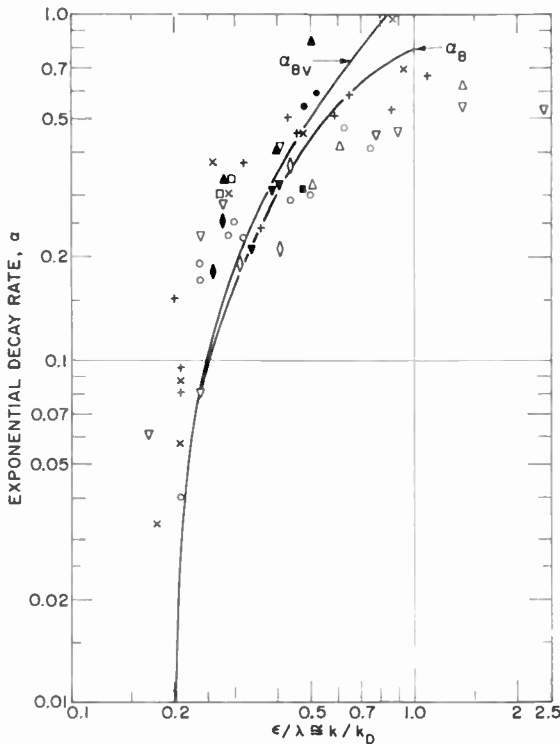


Fig. 5—Experimentally determined exponential damping rates α as a function ϵ/λ and k/k_D (symbols refer to different beam conditions). Reproduced from Reference (9).

dicted by α_{BV}, but, as described,[9] large damping curves are difficult to measure and tend to be on the low side. In addition the normalization equations used a measured interference length which would underestimate α.

To conclude, the rates follow in the main the predictions of the BBV theory. Neither beam irregularities nor transverse effects appear to play any appreciable role. Figure 5 exhibits a remarkable agreement with the theoretical decay rates. Since the BBV theory should, from our earlier considerations, predict the Landau damping rates in electron beams, we consider the data in good agreement with that predicted by Landau. The original designation[1] of the author that this

decay of space-charge wave was Landau damping now appears to have quantitative justification.

Summary and Conclusions

The bulk of the experiments described in Reference (9) showed behavior that can be predicted by Landau damping in electron beams, as developed in the BBV theory. The experiments are mainly in the region of plasma frequency small compared to the modulating frequency. Gunshor[8] and Mihran,[2] on the other hand, describe current-modulation experiments with relatively larger plasma frequencies. They distinguish between a ballistic phase mixing ($\sin\theta/\theta$ damping) and the exponential-type Landau damping in light of the fact that Landau did not explicitly describe the ballistic case. The author does not wish to argue the semantics of the designation, but classifies Landau damping as multivelocity effects causing damping by nondissipative forces. The measured velocity distribution of both Gunshor and Caulton show both a low-velocity tail and a rounded or uniform region. The Bloom and Vural papers[4,5] contain implicitly a description of such a $\sin\theta/\theta$ decay for a current modulated beam with a rectangular distribution.

Gunshor[8] describes such a decay; we suggest it is due to the use of a current or density-modulated beam coupled with a large rounded or flat portion of the velocity spread, and a relatively small number of electron contributing the Maxwellian-type low-energy tail. Thus their so-called ballistic decay associated with the flat portion appears to override the exponential decay due to the low-energy tail. The current-excited beam theory of Bloom and Vural,[5] when the proper initial conditions are used, describes a decay similar to the "ballistic decay." This perhaps explains their successful comparison with Mihran's experiment,[2] although his experimental beam did not have the cut-off Maxwellian distribution they postulated.

This paper describes experiments on a velocity-modulated beam. Berghammer[3] predicts that a rectangular distribution on a velocity-modulated beam would yield no damping, and indeed the experiments described indicate damping solely of an exponential-decay nature, suggesting that any rectangular-type parts of the distribution did not influence the decay.

The equivalence of the BBV results for an electron beam and Landau damping for a stationary plasma has been shown for the small ϵ/λ or k/k_D limit. If the BBV theory holds for larger ϵ/λ (it should be limited only to small ϵ), then Landau damping predictions most closely applicable to an electron beam are valid. Over 50 experiments of plasma

oscillations in electron beams showing wave decay in which the data could be analyzed and the results exhibited in the framework of the BBV and Landau theories were described in Reference (9). While the individual experiments are of limited accuracy, the bulk of the results quantitatively agree well with the predictions of the theory over a wide range of damping parameters. The results then constitute a quantitative experimental verification of the existence of Landau damping in electron beams.

ACKNOWLEDGMENTS

The author wishes to acknowledge fruitful discussions with many of his RCA colleagues, particularly with S. Bloom, L. S. Nergaard, and H. Sobol, and also T. G. Mihran of General Electric and R. L. Gunshor of Rensselaer Polytechnic Institute.

APPENDIX I—EVALUATION OF DAMPING RATES FROM R-F POWER VERSUS DISTANCE CURVES

The BBV publications did not derive the damping rates as an explicit function of ϵ/λ. Figure 6 shows the r-f power interference pattern of a slow wave and fast wave with one of the waves decaying at an exponential rate α'. The dotted lines are loci of the maxima and minima. The difference term, max–min, at a given z will vary as exp $\{-\alpha' z\}$, as may be seen from Figure 6. Semi-log plots of max–min as a function of z yield straight lines and their slopes yield α'. Such plots were made from Berghammer's[3] Figures 3, 4, and 5 for ϵ/λ of 0.25, 0.5 and 1. These points are shown in Figure 4.

Bloom and Vural[4] show current damping patterns for ϵ/λ of 0.3 to 0.9. Again semi-log plots of maxima-minima yield straight lines and the slopes yield α. These rates are shown in Figure 4 along with an extrapolated decay rate at $\epsilon/\lambda = 0.2$. The spatial decay rate is readily converted to the temporal decay one represented as ω_i/ω_p. In particular $\lambda\theta$, the normalized distance in BBV, is $(\omega_p/\omega_m)(\omega_m/u_s)\, z = \omega_p t$.

APPENDIX II—DAMPING RATES WITH HIGH- AND LOW-VELOCITY TAILS

Jackson[10] indicates that the Landau damping term for stationary plasmas is the imaginary part of the plasma frequency ω_i. He derives for *small damping* our Equation (2).

$$\omega_i \cong \frac{\pi}{2}\,\omega_r \left(\frac{\omega_p}{k}\right)^2 \frac{1}{n_o} \left(\frac{\partial f_o}{\partial u}\right)_{u\,=\,\omega_r/k} \left(1 - \frac{k}{\omega_r}\frac{d\omega_r}{dk}\right) \tag{27}$$

where ω_r is the real part of the plasma frequency, n_o is the number of particles, and k is the phase constant (invarient to a Galilean transformation). This expression, although derived from a stationary plasma, is invarient[17] for a Galilean transformation, and we can apply

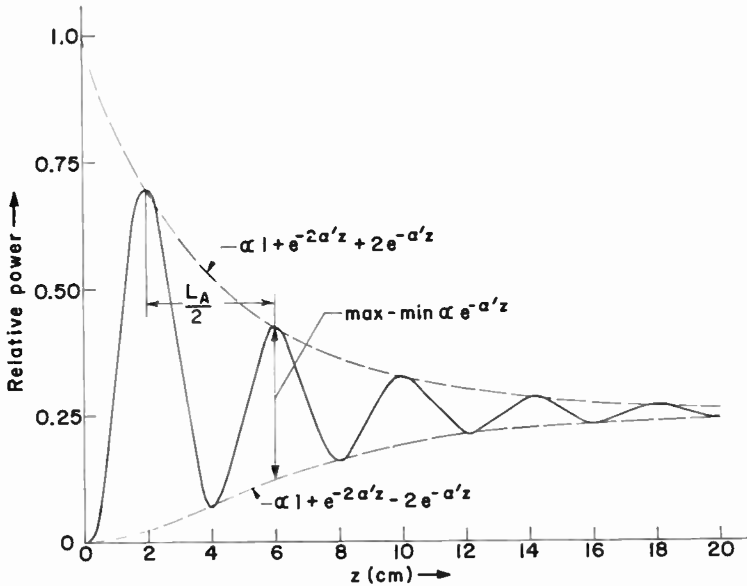


Fig. 6—Computer relative power versus distance z in space-charge waves (current amplitude of one wave decays as $e^{-\alpha' z}$; in References (3) and (5) the abscissa is $\lambda\theta$ and the wave decays as $e^{-\alpha\lambda\theta}$).

it to a moving electron beam. In doing so, one must use the quantities appropriate to the system. For the laboratory electron-beam system

$$\left(1-\frac{k}{\omega_r}\frac{d\omega_r}{dk}\right)=1-\frac{v_g}{v_p} \tag{28}$$

where v_p is the phase velocity and v_g is the group velocity.

It will be demonstrated that the damping expression given by Equation (27) will yield approximately the same results for two distributions of $f_o(u)$. One is a low-velocity tail that is approximately Maxwellian:

---

[17] J. M. Dawson, "On Landau Damping," *Phys. Fluids*, Vol. 4, p. 869, July 1961.

$$f_s = \exp\{-\alpha\,(u_c^2 - u^2),\ 0 < u < u_c \tag{28}$$

where u is the velocity of the individual electrons in the beam,

u_c is the velocity of the fastest electron $u_c^2 = (2\,e/m)\,v_o$,

$\alpha = m/(2KT)$

V_o is the drift potential,

subscript s refers to slow-velocity distribution.

This distribution is similar to that encountered in practice. The other distribution f_f is given by the cut-off Maxwellian distribution with a high-velocity tail (Equation (8)) written as

$$f_f = \exp\{-\alpha\,(u^2 - u_c^2)\},\ u > u_c\,, \tag{29}$$

where the subscript f refers to the fast-velocity distribution and this is the distribution for which the damping rates are derived in the BBV theory. Here u_c is the same as in Equation (28), but refers now to the slowest electrons. In examining how the damping rates of the two distributions compare, the major terms of concern are

$$\left(\frac{\partial f_f}{\partial u}\right)_{u\,=\,v_{pf}}\left(1 - \frac{v_{gf}}{v_{pf}}\right)_{\text{fast tail}} \quad\text{and}\quad \left(\frac{\partial f_s}{\partial u}\right)_{u\,=\,v_{ps}}\left(1 - \frac{v_{gs}}{v_{ps}}\right)_{\text{slow tail}}$$

In the limit of small damping we use[11]

$$v_{pf} = \frac{v_{of}}{1 - \dfrac{\omega_r}{\omega_m}}\,,\ v_{ps} = \frac{v_{os}}{1 + \dfrac{\omega_r}{\omega_m}}\,,$$

and the approximations

$$v_{of} = v_o\,(1 + \epsilon + \ldots)\,,$$

$$v_{os} = v_o\,(1 - \epsilon + \ldots)\,,$$

$$v_{gf,s} \cong u_c\,.$$

With these approximations, remembering that ϵ and $(\omega_r/\omega_m = \lambda)$ are a few per cent or less, and neglecting second-order terms,

$$\left(\frac{\partial f}{\partial u}\right)_{u=v_g}\left(1-\frac{v_g}{v_p}\right) \cong -2\,\alpha\, u_c(1+\epsilon+\lambda)\,\exp\{-\alpha\, u_c^2\,(2\,\epsilon+2\,\lambda)\} \quad \text{(fast)}$$

$$\cong -2\,\alpha\, u_c(1-\epsilon-\lambda)\,\exp\{-\alpha\, u_c^2\,(2\,\epsilon+2\,\lambda)\} \quad \text{(slow)} \qquad (31)$$

We conclude that (1) the sign of ω_i is negative for both distributions as expected, (2) the exponential term, which is the principle factor, is the same for both distributions, and (3) there is slight difference in the damping from the other parenthetical terms that would be insignificant for (a) small ϵ or small damping and (b) for a small ω_r/ω_m (which is the case in these experiments). As the damping term, Equation (27), is valid for small damping only, it is concluded that there is no significant difference between the damping terms of a slow or fast tail half-Maxwellian distribution. That this holds for large damping remains to be proved.

SONIC FILM MEMORY *†

By

H. WEINSTEIN‡, L. ONYSHKEVYCH, K. KARSTAD, and R. SHAHBENDER

RCA Laboratories
Princeton, New Jersey

Summary—In the sonic-film memory, thin magnetic films and scanning strain waves are combined to realize a nonvolatile block-oriented random-access memory. The interaction of strain waves with magnetic thresholds is used to obtain serial accessing and nondestructive read signals. Experimental data showing coincidence selection and nondestructive read operation is presented for a variety of geometries and strain-pulse characteristics. For high bit densities and transfer rates, strain pulses of 10 nanoseconds duration, and 10^{-4} amplitude are required. Transducers capable of generating these pulses have been developed and are described. Experimental data are presented. A reduction in the required strain amplitude by a factor of 5 is realized by using Fe-Ni-Co films. The results of magnetic materials investigation leading to the new composition are presented.

INTRODUCTION

THE SONIC film memory represents a novel approach to the storage of digital information. Thin magnetic films and scanning strain waves are combined to realize a memory in which information is stored serially. The remanent property of magnetic films is used for nonvolatile storage. The effect of strain waves on magnetic films is used to obtain serial accessing. This effect is also used to derive a nondestructive read signal for interrogation.

The sonic film memory represents one possible realization of a block-oriented random-access memory[1] (BORAM). In such a memory, information is stored in blocks with random access to blocks being accomplished at electronic speed. Information in a block is retrieved serially at bit rates compatible with computer speeds.

* The work described in this paper is partially supported by the USAERL under contract numbers DA36-039-AMC-03248(E) and DA28-043-AMC-01392(E).

† This material was presented at the 1966 Fall Joint Computer Conference, San Francisco, Calif.

‡ Now at the Xerox Corporation, Webster, N. Y.

[1] B. H. Gray, D. R. Hadden, Jr., and D. Haratz, "Block Oriented Random Access Memory," *Proc. National Symposium on The Import of Batch Fabrication on Future Computers*, Los Angeles, Apr. 1965.

System Description

The simplest configuration for a sonic film memory block[2,3] consists of a glass tube with a magnetic film coating. A conductor threads the tube and an ultrasonic transducer and an ultrasonic absorbing termination are attached to the ends, as shown in Figure 1. The magnetic coating has square-loop properties and is originally saturated circumferentially in one direction along the entire length. To enter information serially into the block, a strain wave is launched along the tube and propagates from the transducer end to the termination, where it is completely absorbed. As the wave propagates along the tube, it reduces the switching threshold of the strained region. A pulse of current applied to the conductor linking the magnetic coating will cause flux reversal only in the strained region, permitting the selective writing of information.

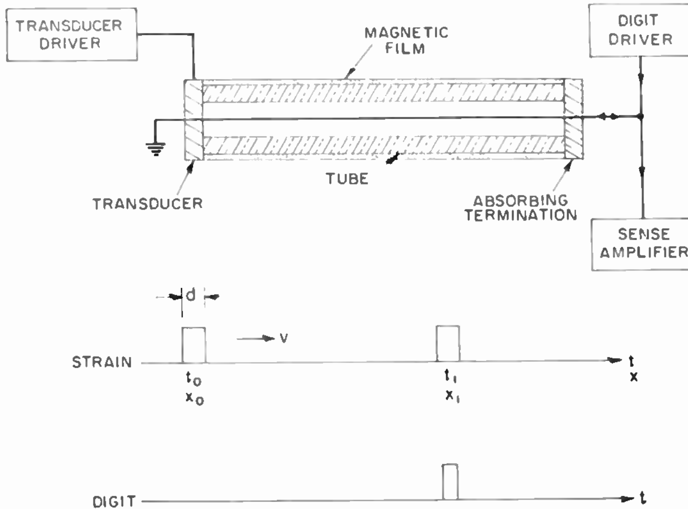


Fig.1—Simple sonic film memory configuration.

Nondestructive serial retrieval of the stored information is effected by propagating a strain wave along the rod. This produces a reversible change in the magnetization of a strained region. This change induces in the linking conductor a voltage whose polarity depends on the remanent state of the flux corresponding to the stored information.

---

[2] J. A. Rajchman, *RCA Technical Notes*, RCA TN No. 346 (1959).

[3] J. W. Gratian and Freytag, "Ultrasonic Approach to Data Storage," *Electronics*, May 4, 1964, p. 67.

Figure 1 shows that the region in space occupied by the strain wave determines the dimensions of a "bit." The spatial dimension d, shown in Figure 1, is given by $d = vt$, where v is the velocity of sound, and t is the time duration of the strain wave.

To enter information, it is necessary that the strained magnetic region complete its switching in a fraction of the time duration of the strain pulse. For high bit densities, say in excess of 50 bits to the inch, this implies a switching time of the order of 10 nanoseconds.

Assuming that the strain wave causes a reduction in the magnetic threshold of only a factor of 2, it is clear that switching must be accomplished with relatively modest overdrives. The requirement of high-speed switching with relatively modest overdrives rules out the possibility of using isotropic magnetic materials for the storage medium. However, thin magnetic films with uniaxial anisotropy have the required property of very fast switching (of the order of 2 nanoseconds) with drives of the order of the anisotropy field.[4]

For the memory configuration shown in Figure 1, and high bit densities, it is necessary that the magnetic coating have uniaxial anisotropy and be capable of fast switching. Such coatings with a circumferential easy axis and an axial hard axis are in use for random-access memories.[5]

STRAIN DEPENDENCE OF MAGNETIC CHARACTERISTICS OF FILMS WITH UNIAXIAL ANISOTROPY

The gross magnetic properties of a planar film with uniaxial anisotropy may be described by specifying the direction of the anisotropy (easy axis), the magnitude H_k of the anisotropy field, and H_c the easy-axis coercive field. The dependence of the anisotropy direction and anisotropy field on uniaxial strains applied at an arbitrary angle with respect to the easy axis may be derived analytically.[6-8] The results of these derivations are

---

[4] H. Weinstein, *An Investigation of Serial, Nonvolatile Computer Memories Based on a Magnetoelastic Interaction in Ferromagnetic Storage Media*, Ph.D. dissertation, Polytechnic Inst. of Brooklyn, E. E. Dept., 1965.

[5] T. R. Long, "Electrodeposited Memory Elements for a Nondestructive Memory," *Jour. Appl. Phys.*, Vol. 31, p. 123s (1960).

[6] T. S. Crowther, "Angular and Magnitude Dispersion of the Anisotrophy in Magnetic Films," *Jour. Appl. Phys.*, Vol. 34, p. 580 (1963).

[7] E. N. Mitchell, G. I. Lykken, and G. D. Babcock, "Compositional and Angular Dependence of the Magnetostriction of Thin Iron-Nickel Films," *Jour. Appl. Phys.*, Vol. 34, p. 715 (1963).

[8] H. L. Pinch and A. A. Pinto, "Stress Effects in Evaporated Permalloy Films," *Jour. Appl. Phys.*, Vol. 35, p. 828, (1964).

$$\tan 2\theta_r = K \frac{\sin 2\phi}{1 + K \cos 2\phi}$$

$$h_k = H_k/H_k^0 = (1 + K^2 + 2K \cos 2\phi)^{1/2}$$

$$= \sin 2\phi/\sin 2(\phi - \theta_r)$$

where ϕ = angle between the strain axis and the unstrained easy axis,

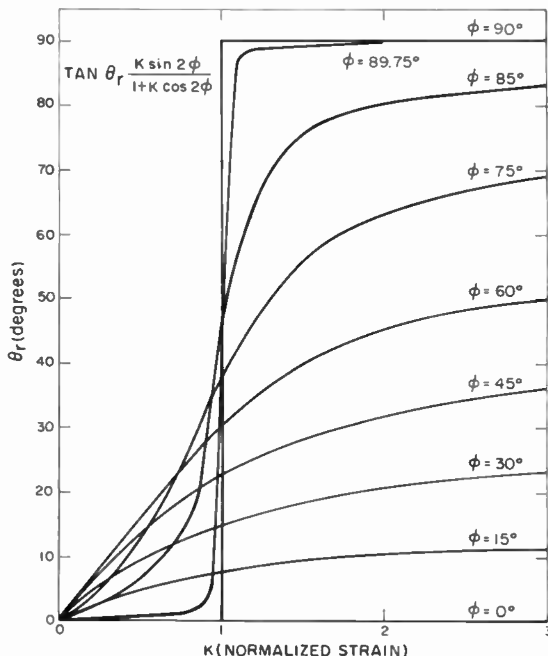


Fig. 2—Axis rotation angle θ_r versus strain parameter K.

θ_r = the angle through which the strain rotates the easy axis (referred to the unstrained easy axis),

H_k^0 = the strain-free anisotropy field,

H_k = the effective anisotropy field in the strained film,

K = strain parameter $= \frac{S}{H_k^0} \frac{\Delta H_k}{\Delta S}$ for $\phi = 0$ or $\pi/2$, and

S = strain.

Figures 2 and 3 are plots of the above equations.[8] In Figure 2, θ_r is plotted as a function of K with ϕ as a parameter. In Figure 3, H_k is plotted versus K with ϕ as a parameter.

Inspection of Figure 2 shows that if the strain is applied along the easy axis ($\phi = 0$), the anisotropy direction is unchanged for all values of strain (tensile or compressive). On the other hand, if the strain is applied at right angles to the easy axis ($\phi = \pi/2$), the anisotropy direction changes discontinuously at $K = 1$ from $\theta_r = 0$ to $\theta_r = \pi/2$.

Inspection of Figure 3 shows that H_k can be increased or decreased by application of strain. For a positive magnetostriction coefficient and

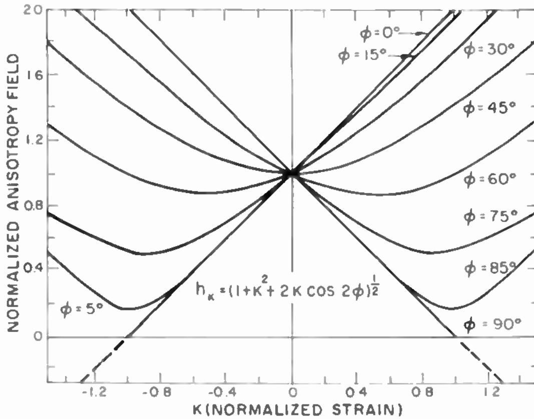


Fig. 3—Normalized effective anistropy field versus strain parameter K.

a tensile strain, K is positive (a compressive strain corresponds to K negative). For this case the normalized anisotropy field H_k is reduced by the application of a compressive strain along the easy axis ($\phi = 0$) or a tensile strain along the hard axis ($\phi = \pi/2$). Figure 3 shows that for values of $K > 1$ and $\phi = \pi/2$, $H_k < 0$. This is a result of the discontinuous rotation of the easy axis as indicated in Figure 2.

Experimental data on planar thin films,[8] and films with cylindrical symmetry[9] are essentially in agreement with these theoretical predictions.

For highly magnetostrictive iron–nickel films the experimental data indicates that a strain of $\sim 10^{-4}$ is required to produce discontinuous rotation of the easy axis. Similarly, a reduction in the anisotropy field by a factor of 2 is realized with a strain of $\sim 10^{-4}$. To accomplish sequential selection in a memory block, either the rotation of the easy axis with strain, or the reduction of the anisotropy field with strain, or

[9] H. Weinstein, "Static and Dynamic Stress Effects in Cylindrical Ferromagnetic Films," 11th Annual Conference on Magnetism and Magnetic Materials, San Francisco, Nov., 1965.

a combination of both effects may be used. In either case, the required strain is of the order of 10^{-4}, and it is necessary to generate this strain at the film. Use of magnetic compositions with higher magnetostriction permits a reduction in the required strain amplitude.

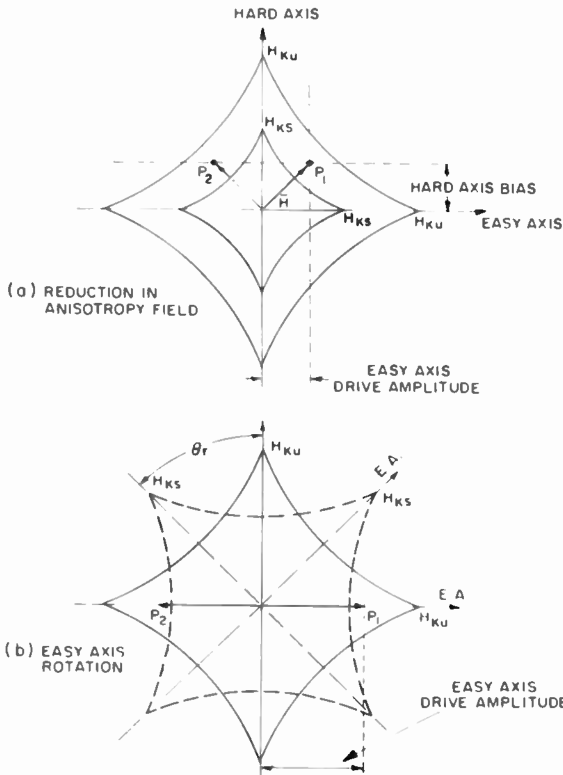


Fig. 4—The switching astroid.

STRAIN SWITCHING OF THIN MAGNETIC FILMS

The switching thresholds for thin magnetic film are given by the well-known astroid shown in Figure 4. For an unstrained element the thresholds are determined by the value H_{ku} and the outside astroid in Figure 4a. For a strained element, the threshold is given by H_{ks} shown by the inside astroid in Figure 4a. This assumes that the strain is applied along the easy axis or along the hard axis and is of a magnitude so as not to produce discontinuous rotation. Consider the points P1 and P2 in Figure 4a. It is clear that a hard-axis bias in combination with either a positive or a negative easy-axis drive will cause

switching of a strained element but will not cause switching of an unstrained element. Thus a coincidence of an easy-axis drive with a strain pulse will cause switching at the local strained region.

Alternatively, the rotation of the easy axis due to the application of strain may be utilized to enter information into the memory. In this case, the strain rotates the easy axis and an easy-axis drive completes switching of the strained region. This is illustrated by the astroids shown in Figure 4b. Bipolar easy-axis drives are required to produce both remanent states. In an actual film, a combination of both rotation and reduction in the anisotropy field is probably the cause of switching.

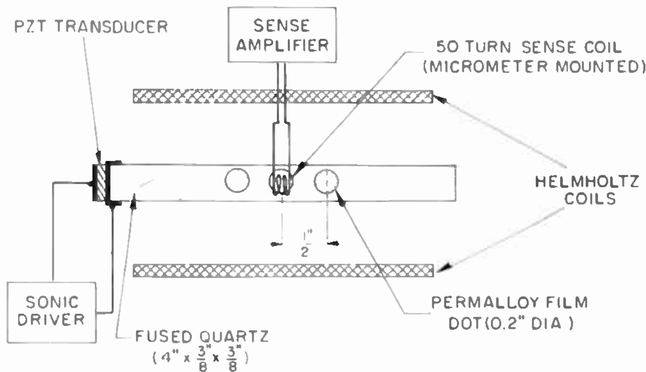


Fig. 5—Memory-device geometry.

Memory-Cell Test Data

Static Magnetic Fields

The storage system selected for initial testing consists of thin permalloy films deposited on a fused quartz substrate. The substrate measures $4 \times 3/8 \times 3/8$ inches and has a PZT (lead zirconate—lead titanate) transducer bonded to one end surface as shown in Figure 5. The PZT transducer operating in the longitudinal mode is used to generate strain pulses in the substrate. This is accomplished by exciting the transducer at its resonant frequency of 4 MHz from an external generator with a modulated sine wave voltage.

The thin-film storage elements have a composition of 60:40 nickel–iron and a diameter of 200 mils. The film spots are spaced 0.5 inch apart to eliminate magnetic coupling between the spots from interfering with the measurements. The hysteresis loops measured under static strains for samples of this composition indicated a uniaxial anisotropy together with a relatively high strain sensitivity. The sub-

strate with the magnetic films is inserted in the mounting jig shown in Figure 6.

For the device shown in Figure 5, the strain pulse propagates in a direction along the hard axis of the magnetic films. Helmholtz coils provide bipolar magnetic fields of approximately 5 oersted in the easy direction during writing and are used to establish the original direction of

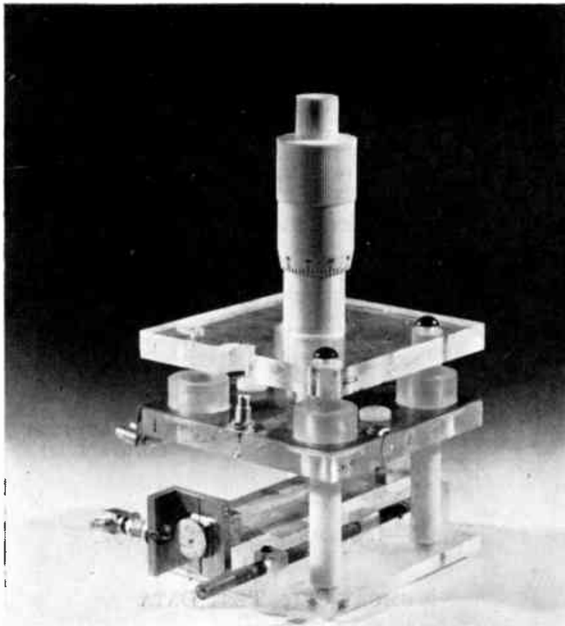


Fig. 6—Sonic film memory rod mounted on drive jig.

magnetization of the films. A 50-turn coil is positioned close to a film element and is used to detect the output sense signal. A strain pulse packet is generated by the transducer and sequentially scans the magnetic films. Figure 7 shows the voltage applied to the transducer. The first signal is the voltage applied by the generator to the transducer, and the second wave form is due to the strain wave being reflected from the far end of the quartz substrate and impinging on the transducer. The lower trace in Figure 1 shows the wave form obtained from the sense coil. This wave form is expanded in Figure 8 for both binary remanent states in the films. As can be seen in Figure 8, the sense signals are bipolar in nature. They are truly nondestructive for limited strain amplitudes.

The magnitude of the sense signal was found to be linearly de-

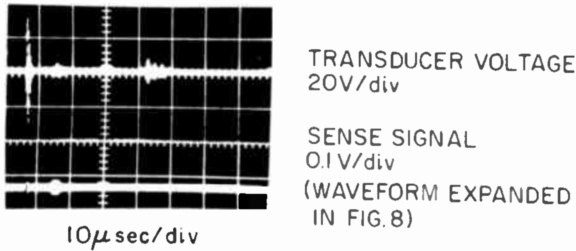


Fig. 7—Sonic drive and sense signal.

pendent on both the strain amplitude and the diameter of the film spot. When a static easy-axis magnetic field is maintained during the sense operation, the magnitude of the sense signal diminishes. For a field of 7 oersted, the sense signal is reduced to zero.

A threshold value exists for the static magnetic field needed to switch a film with a given strain amplitude. The higher the amplitude of the strain pulse, the lower the threshold value for causing magnetic reversal. For transducer excitation of 100 volts peak value, the magnetic threshold is 2.8 oersted. For a sonic drive amplitude of 50 volts peak value, no change in magnetization is observed when the magnetic field is reversed. This demonstrates that a coincidence of a magnetic field and strain pulse is required to cause switching in the magnetic field.

To further demonstrate the need for coincidence between strain pulses and magnetic field pulses to produce switching, dynamic "Write–Read Disturb" testing was undertaken.

Dynamic Testing

For dynamic testing the sense coil of Figure 5 is used to apply the magnetic bit field. A common clock provides synchronization between the sonic and magnetic pulses. The sonic drive program consists of a

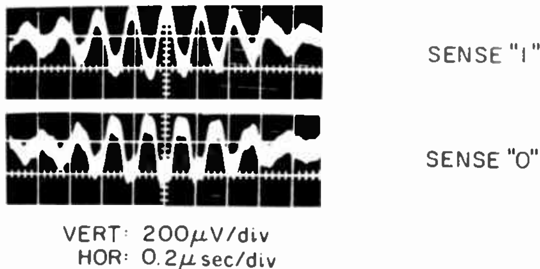


Fig. 8—Expanded sense signals.

doublet of sinusoidal bursts as shown in the timing diagram of Figure 9. The first burst is used to write information into a memory element. The second burst is used to read out the information. The magnetic field pulse is applied via the sense coil in coincidence with the strain burst arriving at the storage element. The upper trace in Figure 10(a)

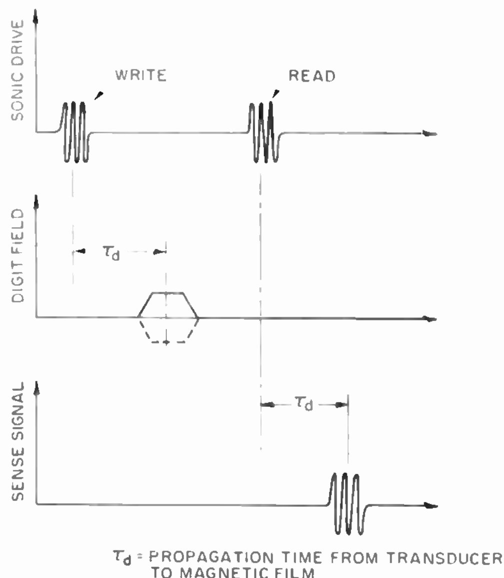


Fig. 9—Timing sequence for dynamic testing.

shows the voltage wave form applied to the transducer for a write–read sequence. The lower trace illustrates bipolar magnetic digit drives (one and zero) timed to coincide with the arrival of the write strain pulse at the test storage location. Figure 10(b) illustrates the disturb condition. Only magnetic energy is applied to the memory cell under test without a write strain pulse. Figures 10(c) and 10(d) show the sense signals corresponding to the write and disturb conditions, respectively. For Figure 10(c) the sense signals are bipolar corresponding to the binary state. Figure 10(d) shows the sense signals obtained by disturbing the binary one state by both positive and negative digit pulses. As can be seen, the digit pulses produce very little disturb effects.

For this set of dynamic tests, distinct thresholds were observed for both the strain and magnetic drive amplitudes. However, the margins of operation for both the strain and magnetic field pulses were narrower than in the case of static magnetic testing.

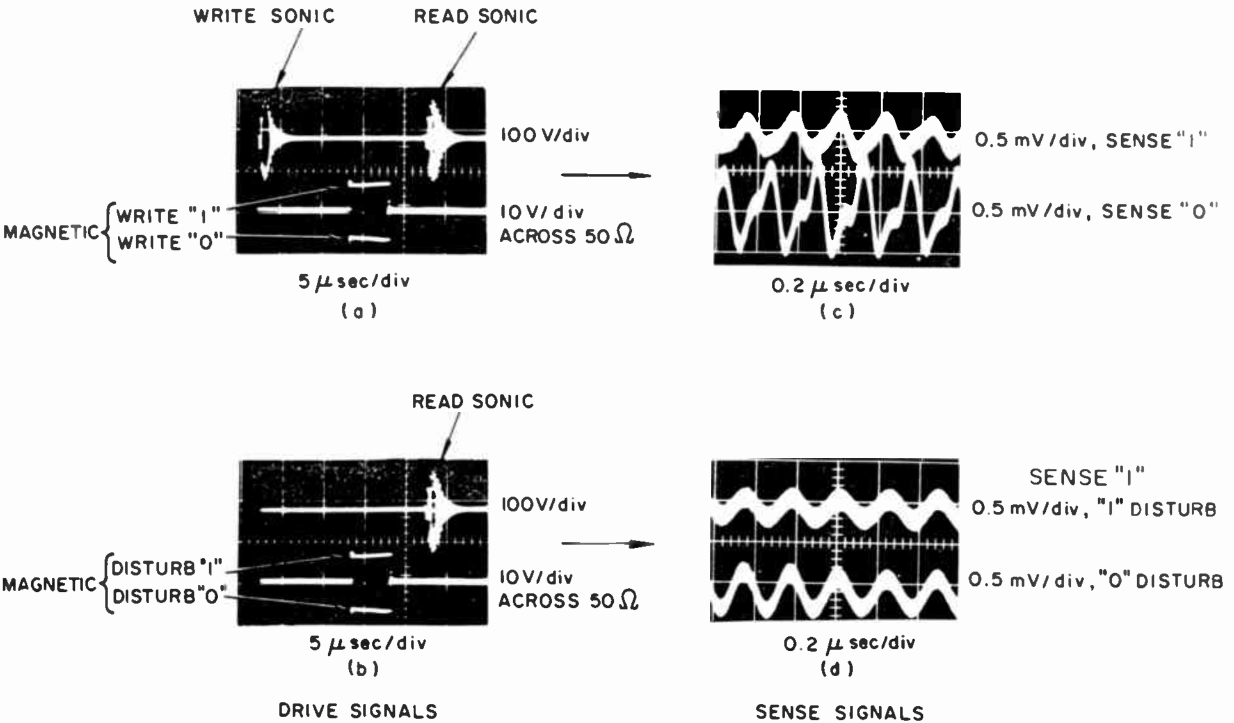


Fig. 10—Memory-cell operation.

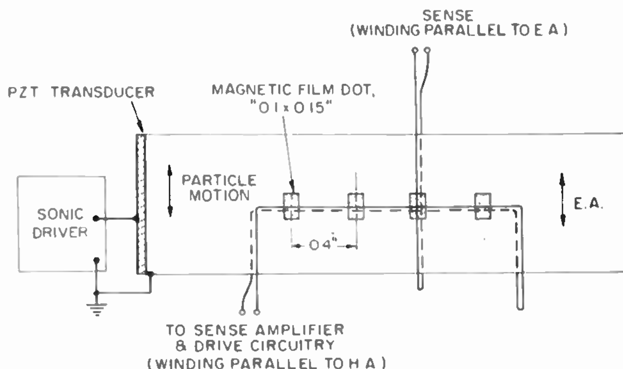


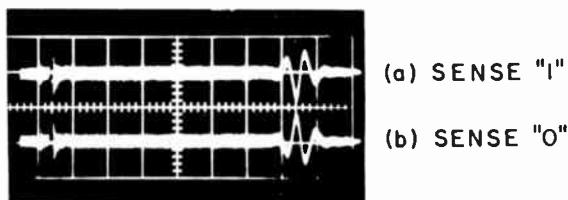
Fig. 11—Planar film storage array.

Unipolar Strain Pulse Testing

For this set of tests, films with an iron–nickel–cobalt composition having very high magnetostriction are used. The elements, with a rectangular shape, measure 0.1 × 0.15 inch and are approximately 1000 Å thick. These films are deposited on conventional microscope glass slides measuring 1 × 3 × 0.04 inches. A shear mode PZT transducer, resonating at 1.5 MHz, is bonded to one end of the glass slide; a fine layer of viscous material is used for damping ultrasonic reflections. Figure 11 shows the geometrical arrangement of the magnetic films, direction of the easy axis, and the direction of particle motion in the shear mode wave.

The glass substrate is sandwiched between two copperclad etched circuit boards that form the sense and drive windings for the magnetic films. Unipolar pulses from an external generator are applied to the transducer to generate strain pulses in the substrate. Excitation ranges from 30 to 100 volts and is adjusted in width for optimum transducer response.

Figure 12 shows the sense signal (sense winding parallel to easy



0.5 mV/div., 2 μsec./div.

Fig. 12—Sensing a single bit.

axis) generated when a given storage location is switched into one of the two remanent states. This is accomplished by the coincidence of a strain pulse generated by the transducer and a magnetic field pulse generated by the drive winding. As can be seen, the sense signals are bipolar and of an amplitude of 0.5 millivolt.

The magnitude of the sense output may be enhanced by applying a constant hard-axis bias during the sense operation. For the device configuration of Figure 11, a hard-axis bias is applied by a pair of Helmholtz coils. The output sense signals are shown in Figure 13. In the absence of a hard-axis bias, the sense signals induced in a winding parallel to the hard axis are negligible.

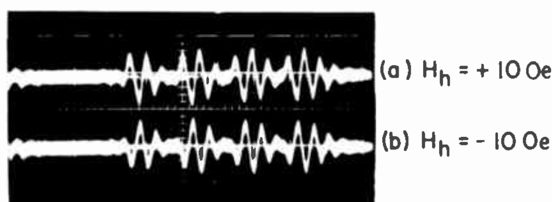


Fig. 13—Sense signal from a 4-element storage array.

STRAIN-WAVE GENERATION

For operations of a high-density sonic film memory, unipolar strain pulses with a width of approximately 10 nanoseconds and a strain amplitude no greater than 10^{-4} must be radiated into a fused quartz substrate. The most practical method for generating such pulses is by means of piezoelectric transducers.

Of the many piezoelectric materials, five were selected for investigation—quartz, cadmium sulfide (CdS), lithium metagallate ($LiGaO_2$), antimony sulfur iodide (SbSI), and lead zirconate-lead titanate (PZT). The emphasis was on quartz, which is readily available and capable of fundamental-mode operation up to approximately 100 MHz. PZT is used in the low MHz range because of its high conversion efficiency. The three other materials were investigated because of their great potential. A brief survey is given below of the piezoelectric characteristics of these materials.

Material Characteristics

(1) Quartz

For X-cut quartz vibrating in its thickness mode, the strain gener-

ated[10] by an applied electric field, E, is

$$s = d_{11} \times E$$

where d_{11} is the transmission sensitivity. For quartz $d_{11} = 2.3 \times 10^{-12}$ coulomb/newton.

The strain propagating in a fused quartz substrate may be evaluated by using a transducer as a receiver. The electric field generated by the transducer due to an incident strain pulse is

$$E = g_{11} \times T = g_{11} \times Y \times s$$

where T is the incident stress, Y is Young's modulus for fused quartz, and g_{11} is the reception sensitivity. For quartz,

$$g_{11} = 0.058 \text{ volt-meter/newton.}$$

For an exact determination of strain at a specific location, losses in the bond and the substrate must be taken into account.

To generate a narrow strain pulse, a wide-band transducer is required. The pass band of a transducer is defined almost entirely by the ratio of the acoustic impedance of the transducer to that of the delay medium.[11] Fortunately, the characteristic acoustic impedance of quartz (15.1×10^{-6} kg/m²—s), and fused quartz (13.1×10^{-6} kg/m²—s), are almost equal. A quartz crystal loaded with fused quartz and air backing has a mechanical quality factor $Q_m \approx \pi/2$.

The objective is to have a strain pulse with a rise or fall time τ of the order of 5 nsec. This requires a transducer bandwidth B of 80 MHz ($\tau \times B \approx 0.4$) and a natural-resonance frequency in excess of 60 MHz.

For a 1-mil-thick transducer, the resonance frequency is 100 MHz. To generate a strain of 10^{-4} requires 1100 volts. The emf generated by the transducer operating as a receiver due to an incident strain of 10^{-4} is 11.5 volts.

Quartz has a low dielectric constant ($K = 4.5$); the acoustic losses are extremely small and the capacitive loading is unimportant for the required pulse response.

(2) *Cadmium Sulfide, (CdS)*

CdS has a higher transmission sensitivity than quartz:

---

[10] W. P. Mason, *Physical Acoustics and the Properties of Solids*, Van Nostrand, Princeton, N. J., 1958.

[11] Brockelsby et al, *Ultrasonic Delay Lines*, Iliffe Books, London, 1963.

$$d_{33} = 10.3 \times 10^{-12} \text{ coulomb/newton.}$$

The single-crystal material is difficult to lap down to thin wafers (below 3 mils thick). However, CdS can be evaporated and recrystallized[12] giving an attractive approach for making practical high-frequency transducers. Fundamental-mode operation at 100 MHz requires a film thickness of 17.5 μ, and 175 volts will generate a strain of 10^{-4}.

The characteristic impedance of CdS is 21.5×10^{-6} kg/m^2—s and there is a mismatch to fused quartz representing a 2.3 db loss, which must be added to bond losses caused by imperfect bonding. With evaporation techniques, however, inherently good bonds are expected.

(3) Lithium Metagallate, ($LiGaO_2$)

Relatively little is known about this material.[13] Single crystals are mechanically hard and have a low dielectric constant. The acoustic Q ($> 10^5$) has been measured. The electromechanical coupling coefficient $k_t = 0.25$ and $d_{33} = 7.7 \times 10^{-12}$ coulomb/newton. The compound seems to have most of the attractive qualities of quartz, but is capable of generating three times higher strain for the same drive voltage. The characteristic impedance is 27×10^6 kg/m^2—s and the mismatch of $LiGaO_2$ to fused quartz causes a loss of 3.5 db.

(4) Antimony Sulfur Iodide, (SbSI)

This material has only recently been investigated as a piezoelectric transducer.[14] It is the strongest piezoelectric material known. At 18°C, $d_{33} = 2000 \times 10^{-12}$ coulomb/newton which is three orders of magnitude greater than for quartz. The coupling coefficient k_t ranges from 0.8 to 0.9. Disadvantages are the very high dielectric constant, and the Curie temperature of 22°C.

(5) Lead Zirconate-Lead Titanate, (PZT)

This ceramic compound is readily available and has been investigated thoroughly. Its transmission sensitivity $d_{33} = 260 \times 10^{-12}$ coulomb/newton. A potential of 110 volts across a 20-mil-thick transducer produces a strain of 10^{-4}. The relative dielectric constant is large ($K_3 = 1300$), and makes unipolar pulse operation difficult. The char-

---

[12] N. E. Foster, "Ultra-High Frequency Cadmium-Sulfide Transducers," *IEEE Transactions*, Vol. SU-11, Nov. 1964.

[13] J. P. Remeika and A. A. Ballman, "Flux Growth, Czochralski Growth, and Hydrothermal Synthesis of Lithium Metagallate Single Crystals," *Appl. Phys. Letters*, Vol. 5, No. 9, (1964).

[14] D. Berlincourt, et al, "The Piezoelectric Effect in the Ferroelectric Range in SbSI," *Appl. Phys. Letters*, Feb. 1964.

acteristic impedance is 30×10^6 kg/m^2—s and losses due to mismatch to fused quartz are 4.2 db.

Experimental Data

(1) Quartz

Transducers having thicknesses of 1, 2, and 3 mils were tested in the longitudinal mode of vibration. One-mil thickness corresponds to a resonant frequency of 100 MHz, and is about the upper practical limit for fundamental-mode operation.

For the delay material we used fused quartz because of its low attentuation characteristics. Most of the material used is commercial grade. The end surface of the fused quartz rod is given a conductive coating of gold.

A good bond between the crystal and the metallized rod is difficult to achieve at high frequencies. The bonding agent finally adopted was waterglass (potassium silicate). The tensile strength of the bond is high and its characteristic impedance is close to quartz ($\approx 9 \times 10^{-6}$ kg/m^2—s), although the bond is not of a permanent nature.

We tested the samples to estimate strain magnitude, determine pulse shape, optimize the width of the drive pulse, determine amplitude/frequency response, signal-to-noise ratio, insertion loss, attenuation in bonds and delay medium, and observe extraneous modes. The transducer assemblies are mostly operated single-ended, i.e., the same crystal is used for both transmitting and receiving with the signal reflected from the polished end of the fused quartz.

A figure of merit F may be defined as the ratio of the transducer output voltage when operating as a receiver to the exciting voltage applied to the same transducer. The figure of merit measured for the best samples is:

$$F = 1.4 \times 10^{-3}$$

for a 3-inch-long single-ended sample, and

$$F = 2.5 \times 10^{-3}$$

for a ⅞-inch-long single-ended sample.

These samples were excited with unipolar pulses from a mercury pulse generator. The output voltage is measured across a 50-ohm termination. Up to 1000 volts are applied. The figure of merit values include losses in bonds and in the delay material.

The losses are measured at 100 MHz by observing the exponential

decay of reflected echoes while the transducer is excited with bursts of 100-MHz sine waves. Losses in a bond are 2.45 db, and the intrinsic loss in fused quartz is 0.26 db/cm. Similar values are given in the literature.[15]

Figure 14 illustrates pulse response from 1-mil-thick crystals. The

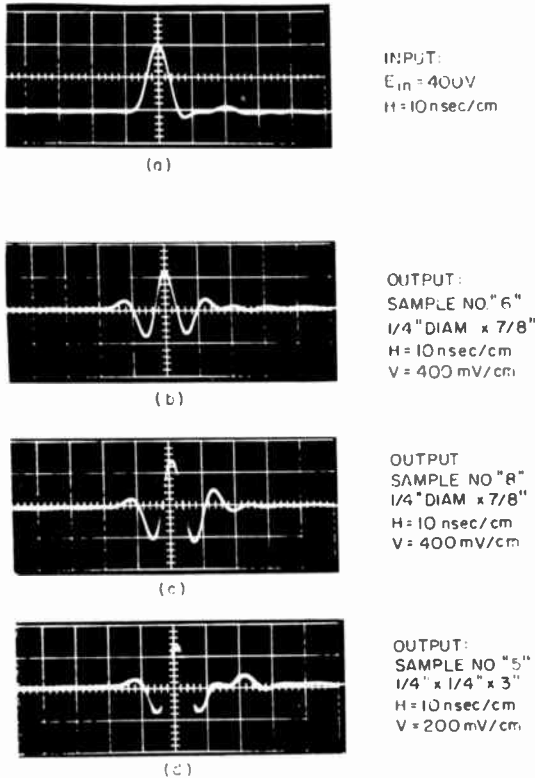


Fig. 14—Output from three different samples of quartz transducers, single-ended.

output amplitude can be maximized by adjusting the width of the input pulse. The optimum input pulse width τ_{in} is related to the transducer resonance frequency f_r by

$$\tau_{in} = \frac{1}{2f_r} = 5 \text{ nsec}$$

We measured $\tau_{in} = 7$ nsec and the output pulse width is $\tau = 6$ nsec.

---

[15] W. P. Mason, *Physical Acoustics*, Academic Press, New York, 1963, Vol. I, part A.

The pulse response is indicative of a wide-band transducer. This is confirmed in the amplitude/frequency response curve shown in Figure 15. The 3-db bandwidth in this curve is 100%. The signal-to-noise ratio of the sample is 67:1. The S/N ratio is defined as the ratio between the first echo and the largest spurious signal following it. The measurement is done with short bursts of sine waves at 100 MHz. While the transducer ideally should vibrate only in its longitudinal mode, some shear modes, mode conversions, and spurious modes will always be generated.

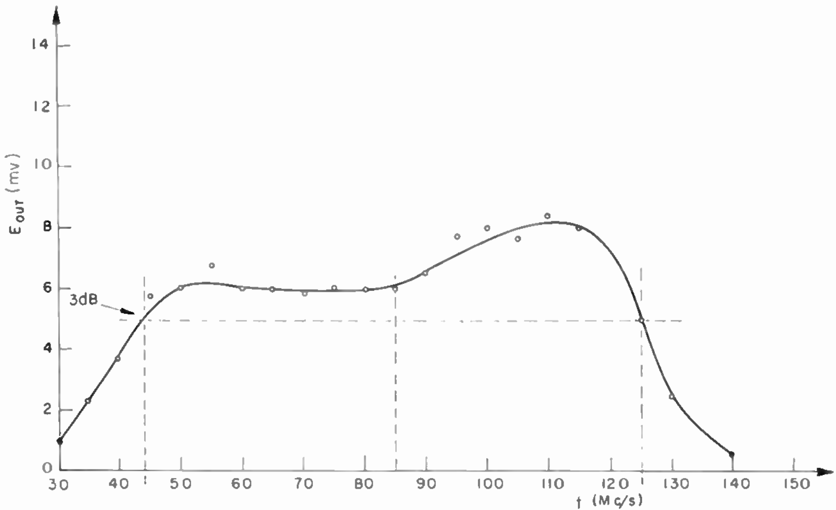


Fig. 15—Frequency response characteristic for 1-mil quartz transducer.

(2) Cadmium Sulfide, (CdS)

High-resistivity films that are piezoelectric active in the 100-MHz and lower frequency range have been produced by an approach based on the de Klerk method.[16]

Figure 16 shows the response from a 1-inch-long sample operated with bursts of sine waves at 100 MHz. The figure of merit (unipolar operation) is $F = 30 \times 10^{-3}$.

(3) $LiGaO_2$, SbSI, and PZT

Lithium metagallate has been used at a natural-resonance frequency

[16] J. de Klerk and E. F. Kelly, "Vapor-Deposited Thin-Film Piezoelectric Transducers," *Rev. Sci. Instr.*, Apr. 1965.

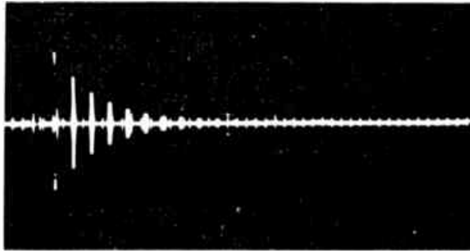


Fig. 16—Echo response from evaporated CdS transducer ($f = 100$ MHz).

of 19 MHz. The figure of merit is about three to four times higher than observed for quartz.

SbSI is a difficult material to cut and polish because of its soft porous nature. One 5-mil-thick sample operated at 23 MHz. However, the figure of merit is far below theoretical expectation.

Experiments with PZT-4 material were carried out for samples as thin as 5 mils and fundamental operation of 17 MHz. The figure of merit for unipolar pulse operation is 130×10^{-3}. This is two orders of magnitude higher than for quartz.

(4) Guided Wave Propagation

To check the feasibility of strain-wave interaction with anisotropic thin magnetic film, a guided-wave structure is employed. In this structure, planar films are evaporated on a standard microscope glass slide ($1 \times 3 \times 0.040$ inch). The glass slide serves as the delay line. As shown in Figure 17, a shear-mode transducer with indicated polariza-

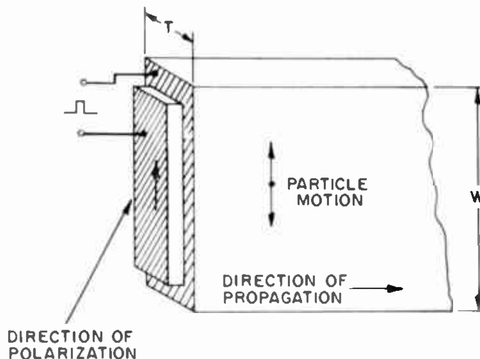


Fig. 17—Strip delay line with shear-mode transducer.

tion is attached to the end of the glass slide. The end surface is first polished and metalized with an evaporated gold electrode. The transducer is bonded to the substrate with silver epoxy.

According to earlier work,[17] an infinite number of width shear modes can propagate in a strip of infinite width. All of these modes are dispersive with cutoff characteristics except for the zero-order mode, which propagates at all frequencies from zero upwards. At frequencies sufficiently low that an acoustical wavelength is greater than twice the strip thickness, only the zero-order mode can propagate and it is completely free from dispersion. From the viewpoint of acoustical transmission alone, the fundamental nondispersive shear mode of propagation offers the most advantageous characteristics. Also the shear-wave velocity is less than the velocity of extensional waves by a factor of about three-fifths, and permits greater bit packing density.

The lowest cutoff frequency of the longest dispersive mode for the chosen substrate geometry is $f_c = 1.9$ MHz. The resonance frequency of the transducer used for testing is chosen as 1.5 MHz, and the acoustical wavelength is 100 mils. Since the strip width is many times the wavelength, the main lobe is narrow in the width direction. Furthermore, the minor surfaces are coated with an absorbing material to lower reflections of energy from secondary lobes.

The echo pattern reflected from the end of the 3-inch-long strip is shown in Figure 18. Figure 18(a) illustrates sine-wave drive, while unipolar-pulse operation is shown in Figure 18(b). The first pulse echo is expanded in Figure 18(c). The optimum width of the input pulse is 330 nsec and the bandwidth of the transducer is 34%.

Magnetic Materials

Requirements

The objective of the investigation of magnetic materials is to find a composition with high sensitivity to strain and low threshold values —H_c or H_k.

In addition, the films must have the following properties: good orientation of the easy axis, good squareness of the hysteresis loop, and low dispersion.

Nickel–iron magnetostrictive alloys have thresholds below 10 oersted, good squareness, and good orientation. The highest strain sensitivity encountered with these materials yields a change in threshold by a factor of 2 for a strain of $\sim 10^{-4}$. For practical operation of a sonic

[17] A. H. Meitzler, "Ultrasonic Delay Lines Using Shear Modes in Strips," *IRE Transactions on Ultrasonic Eng.*, 1960.

memory, a material with a similar change for a strain of 10^{-5} is desirable.

Materials Investigation

To find a desirable film material special instrumentation was constructed for production and static testing of samples as described below.

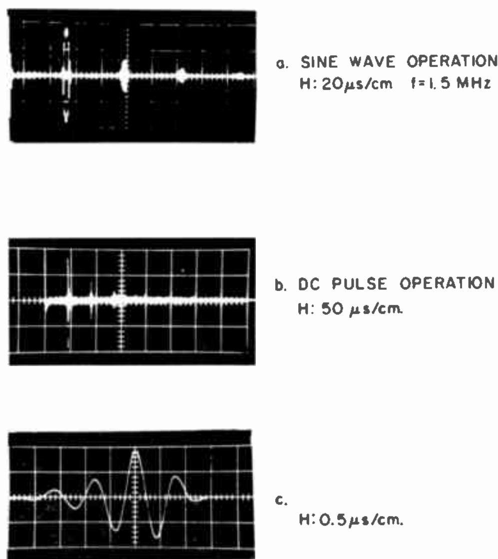


Fig. 18—Echo response for PZT shear-mode transducer on strip delay line $1 \times 3 \times 0.040$ inch.

Fabrication Facilities

For this phase, the films are produced by evaporation. The melt material is evaporated from an alumina crucible by r-f induction heating. Eight substrates can be coated serially in a single run. The vacuum during evaporation is of the order of 10^{-6} torr.

The substrates are heated to 350°C by specially developed individual ceramic heaters. The heat distribution over a slide is rather critical. A thorough investigation shows a maximum variation of only ± 5°C from slide to slide and over a single slide from its edge to the center. The temperature of the slides is closely monitored by thermocouples.

Thickness is measured by a crystal monitor. The reading is differentiated to produce a rate reading; this in turn is fed back to control the power output of the r-f power supply to achieve a constant rate.

Automatic shuttering is used to produce films of very uniform thickness.

The films are evaporated and cooled in a uniform magnetic field of 50 oersted produced by Helmholtz coils.

Both glass and fused silica are used as substrates. Cleaning of the substrates is accomplished by boiling in chromic acid, ultrasonic shaking, and repeated rinsing in acetone, alcohol, and distilled water. Patterns of spots are produced by masking. Figure 19 shows some of the films used in this program. In order to prevent oxidation of some of the more active films, they are coated with plastic spray immediately after opening of the bell jar.

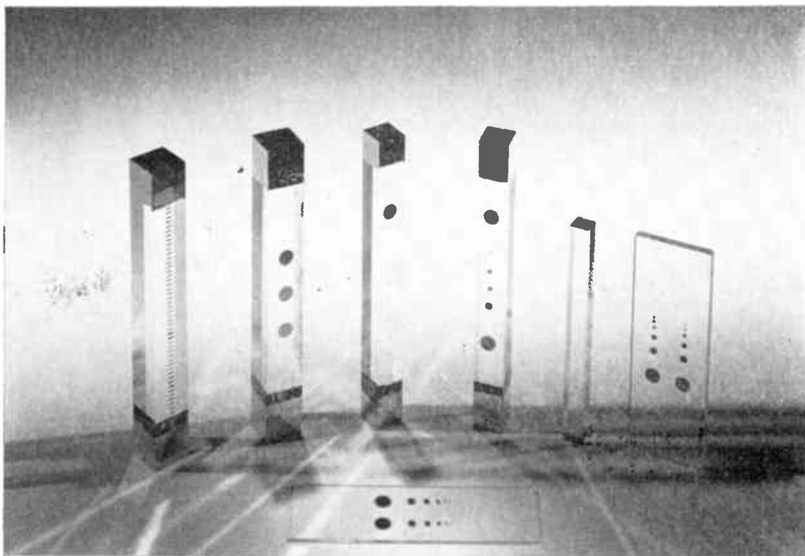


Fig. 19—Typical films produced for testing.

Static Testing Facilities

In order to measure the magnetic parameters of the films under various conditions of static strain, a unique hysteresis-loop tester was constructed, the active part of which is shown in Figure 20.

This instrument permits testing of film spots of various diameters from ½ inch to 20 mils. To accomplish this, sensing heads of various sizes are used. The sensing heads consist of coils, wound in a figure-8 pattern, each half consisting of 200 turns of small wire.

Magnetic drive at 10 kHz is applied to the sample with a maximum of 60 oersted peak-to-peak.

Strain can be applied to the sample—both tension and compression. The substrate is cemented with epoxy into bakelite jaws that are coupled to a pneumatic system. The system can apply static strains in

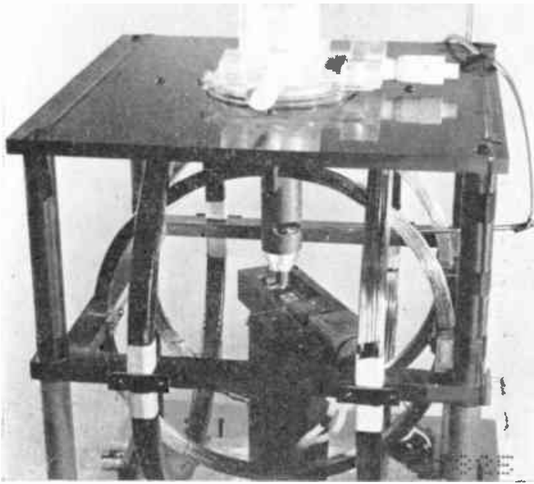


Fig. 20—Apparatus for measuring hysteresis loops.

excess of 10^{-4}. Strain gages are used to monitor the strain induced in the substrate.

Both the sensing coil and the film sample can be rotated and moved laterally with respect to the driving field. A d-c bias field at right angles to the driving field is also provided.

Materials Tested

Figure 21 shows the Ni-Fe-Co melt compositions tested to date. Samples were prepared corresponding to the points indicated in the figure. The films were then tested by fluorescence and wet-chemical-analysis techniques to determine the exact composition. Note that the chart in Figure 21 gives the *melt* compositions rather than the film

compositions. The exact percentages for the eight films evaporated serially from a melt differed by a few percent both from the melt and among themselves. The films were then subjected to testing on the hysteresis loop tester, and, selectively, to x-ray diffraction analysis to determine the crystal structure.

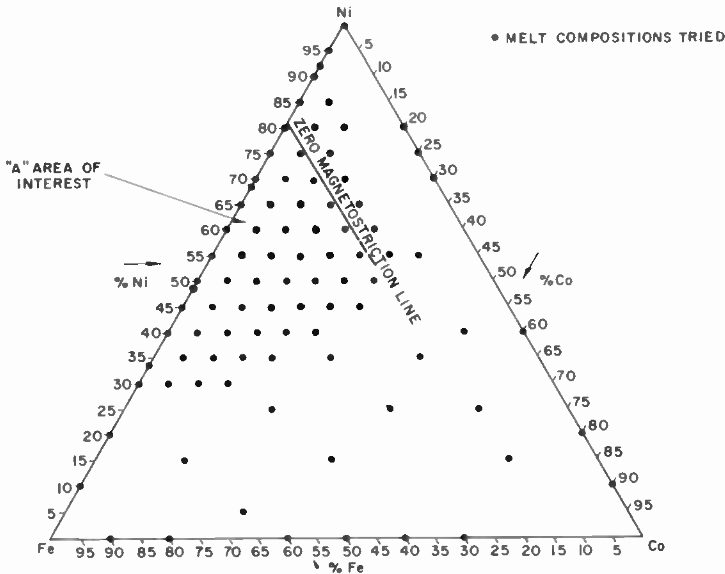


Fig. 21—Nickel–iron–cobalt compositions.

Results

The best results obtained so far are from melts in region A on the chart of Figure 21. Films from melts more iron-rich than that end up with a mixture of a Ni-Fe-Co alloy in a body-centered (α) or face-centered (γ) cubic crystal and precipitated α-iron. More-cobalt-rich films give a mixture of body-centered cubic (α) and hexagonal (ϵ) phases; this causes instability and very high values of coercivity. Nickel-rich films have lower magnetostriction, which actually goes through a null along the line indicated in the chart.

In region A one can obtain very good films, provided a single crystalline phase is achieved. Substrate temperature, surface cleanliness, rate of deposition and other factors determine the film characteristics. Annealing may be used to improve these characteristics.

Figure 22 shows the hysteresis loop of an especially good film obtained from a melt in region A of Figure 21. The photograph shows

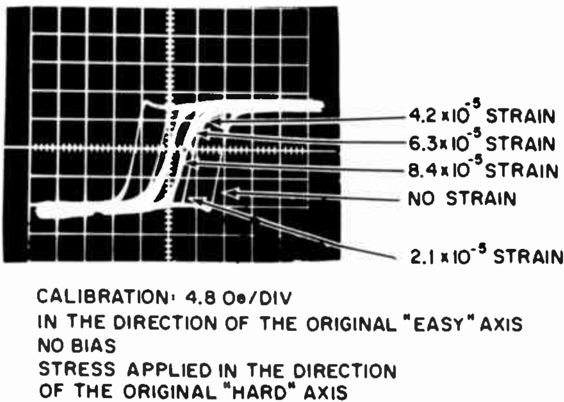


Fig. 22—Stress sensitivity of a magnetic film made from a 25% Fe, 15% Co, 60% Ni melt.

the hysteresis loop in the direction of the original easy axis under various conditions of strain, applied along the original hard-axis direction. The film is well oriented, with a square loop in the easy direction; the unstrained $H_c = 7.7$ oersted, $H_k = 3.8$ oersted. Under strain the easy axis rotates and H_k and H_c change. A strain of 1.6×10^{-5} decreases the switching threshold in the original easy-axis direction by a factor of 2 (see Figure 23). Some films made from this melt composition were used in the dynamic experiments described in the section on memory cell test data.

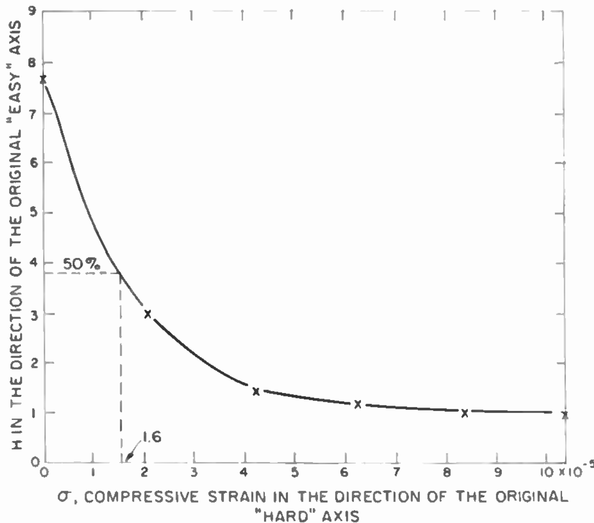


Fig. 23—Coercivity in the easy direction versus strain.

CONCLUSIONS

The experimental data presented show conclusively that rotational switching of thin magnetic films with uniaxial anisotropy may be accomplished by the coincidence of strain pulses and magnetic field pulses. The tested films exhibit thresholds for both strain and magnetic fields. Nondestructive read signals are obtained by subjecting a film element to a strain pulse.

The relative magnitude of the current pulses required to switch a film is compatible with semiconductor circuitry—of the order of 1 ampere. The sense signals derived from elements of 0.1 inch diameter for a strain pulse of 100-nsec rise time are 0.5 millivolt. The magnitude of the sense signal is proportional to the diameter of the film spot and inversely proportional to the rise time of the strain pulse. For smaller film spots, shorter rise times may be used to maintain the sense signals outputs. Thus for a 5-nsec rise time and a spot size of 5 mils, the sense output is expected to be of the order of 0.5 millivolt.

Transducer developments resulted in evaporated CdS films of the required thickness that generated high-amplitude strain pulses of the required speed with modest drive voltages. The developed magnetic materials, with magnetostriction coefficients considerably higher than permalloy, further reduce the transducer drive voltage to levels compatible with semiconductor circuitry.

Magnetic interactions between adjacent bits will undoubtedly be of importance at high bit densities. For random-access film memories, bit densities of the order 50 per inch appear feasible.[18] For the sonic film memory, the number of disturbing magnetic field pulses to which a given bit is subjected does not exceed the number of bits in a block. Thus it is expected that disturb effects will be of smaller magnitude, allowing high bit densities.

Achieving high bit densities and block lengths of several thousand bits is crucial to the economic success of the sonic film memory. Continuing effort is being directed toward achieving this goal.

ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions and guidance of J. A. Rajchman. S. Hotchkiss contributed in the area of mechanical design and was primarily responsible for the design and construction

[18] Q. W. Simkins, "A High Speed Thin-Film Memory—Its Design and Development," *Proc. 1965 Fall Joint Computer Conference*, Vol. 27, part 1, supplement.

of the hysteresis-loop tester described in the paper. D. Leibowitz developed the technique for fabricating CdS film transducers. J. Walentine developed an extremely ingenious technique for reliable bonding of quartz transducers to quartz substrates. A. Monsen and T. Ward were instrumental in many phases of this work and were primarily responsible for device and sample preparation.

Finally, the authors wish to acknowledge the continued interest in the progress of this work by D. Haratz and D. Hadden of the U.S. Army Electronics Research Laboratory, Fort Monmouth, N.J.

BORON NITRIDE AS A DIFFUSION SOURCE FOR SILICON

BY

N. GOLDSMITH, J. OLMSTEAD, AND J. SCOTT, JR.

RCA Electronic Components and Devices
Somerville, N. J.

Summary—This paper describes a boron deposition–diffusion technique based upon the oxidation of boron nitride slices. The technique offers many advantages in conventional two-step (deposition–redistribution) diffusion methods, including elimination of flow-pattern problems, maximum utilization of the deposition furnace, close control of the impurity flow density, use of a single inert gas, and compatibility with other systems and existing equipment. The paper describes the procedure used for the boron nitride system, and shows curves of sheet resistance as a function of deposition time and temperature.

INTRODUCTION

IN SILICON, a two-step diffusion process is normally used when surface impurity concentrations below the solid solubility are required. The two steps are referred to as deposition and redistribution (or drive-in). Ideally, this process consists of a deposition step that produces a thin, highly doped region in the silicon surface. The impurity source and any glass that may have formed on the surface are then removed, and the impurities are redistributed in the silicon by a second heating. Because the number of impurities available during the redistribution step is limited, the surface concentration decreases as the junction depth increases.

In practice, the ideal deposition–redistribution process is only approximated for boron diffusion into silicon, particularly when a steam environment is used during the redistribution step. Because boron is removed from the silicon by a combination of preferential segregation into the growing oxide and volatilization, the total number of diffused impurities becomes a function of both the deposition and redistribution cycles.[1,2]

If steam is not used during the redistribution step, the loss of boron is reduced considerably. As a result, fewer impurities must be intro-

---

[1] J. T. Law, "Surface Properties of Vacuum-Cleaned Silicon," *Jour. Phys. Chem. Solids*, Vol. 14, p. 9, 1960.

[2] A. S. Grove, O. Leistibo, Jr., and C. T. Sah, "Redistribution of Acceptor and Donor Impurities During Thermal Oxidation of Silicon," *Jour. Appl. Phys.*, Vol. 35, p. 2695, Sept. 1964.

duced at the deposition step. In previous boron deposition systems, reducing the total number of impurities, while at the same time maintaining precise doping control and uniformity over the wafer surface, has been an exceedingly difficult problem.

It is apparent that improvements in boron diffusion can be made by improving the control at the deposition step. Prior methods of deposition depend upon the transfer of the doping agent, in its gas phase, from a source that is distant from the heated wafer. The uniformity of the deposition is thus necessarily subject to flow patterns along the length of the furnace tube. Furthermore, the furnace tube itself may act as both a source and a sink for boron during a single deposition cycle, depending upon the vapor concentration of boron being supplied by the distant source.

The limitations of the conventional techniques can be appreciated by examination of Fick's Law,[3] $F = -D \nabla N$, which relates the flow density F of the diffusing impurity to the diffusion constant D and the concentration gradient. This law states that the rate of impurity flow into the silicon is proportional to the impurity-concentration gradient within the silicon. In the early minutes of deposition, the silicon can accept all of the impurity atoms reaching its surface, and the surface concentration is limited by the rate at which the impurity can be supplied. The wafer continues to act as an infinite sink until the concentration gradient decreases to the point where the impurity flow density arriving at the silicon surface (F_s) is greater than the product $D \nabla N$. Until this point is reached, the first sections of the volume in which the diffusion is carried out severely deplete the vapor concentration and, thus, prolong the time before the inequality is fulfilled for regions downstream.* This effect can be reduced by increasing F_s to minimize the period of severe vapor depletion. In practice, F_s is increased by supplying a surplus of dopant impurity in the vapor and by arranging the flow distribution and wafer placement to provide essentially undepleted vapors to all surfaces. The success of the boron tribromide system[4] depends upon application of these principles.

Unfortunately, providing an excess of dopant impurity to reduce

[3] F. M. Smits, "Formation of Junction Structures by Solid-State Diffusion," *Proc. I.R.E.*, Vol. 46, p. 1049, June 1958.

* From wafer to wafer down the length of the furnace tube, allowance must be made for depletion of the vapor stream by both the tube and the boat.

[4] Although this technique is virtually standard throughout the industry, the authors know of no formal published paper describing the system. Some information may be found in, *Integrated Silicon Device Technology, Volume IV—Diffusion*, pp. 110-111, prepared by Research Triangle Institute, Durham, N. C. (AD 603716 available from the Clearinghouse for Federal Scientific and Technical Information, Department of Commerce.)

the nonuniformity from wafer to wafer does not achieve the desired goal, which is to obtain control by limiting the number of impurity atoms introduced. In addition, there are other undesirable side effects. For example, the large excess of boron in the vapor can cause earlier mask failure because more of the masking oxide is converted to a boron glass. Furthermore, the use of high local concentration of boron results in severe damage[5] to the silicon crystal structure.†

The ideal deposition system would have the following features:

(a) equal flow densities to all wafers, with uniform impurity flow density to all portions of a wafer;
(b) an impurity flow density that is easily controlled and independent of wafer temperature;
(c) a positional independence and good space utilization within the furnace.

Obviously, no diffusion technique will satisfy all of these requirements. Although the doped oxide method[6] comes closest, it requires major revisions in equipment and in present diffusion practice. The technique described in this paper, in which the impurity supply is direct rather than indirect, seems to be the next best. This technique, the boron nitride system, can be considered as a modification of the boron box.[7]

THE BORON NITRIDE SYSTEM

The boron nitride system is based upon the oxidation of boron nitride (BN) slices. When oxidized, these slices form B_2O_3 on their surfaces. The oxidized surfaces are then placed between (parallel and close to) the silicon wafers to be diffused, and the assembly is inserted into the heated furnace. In this arrangement, each silicon slice has its own flat, large-area impurity source during the deposition. This system provides the following advantages:

1. Because impurity transport is between two close-spaced, plane parallel surfaces, flow-pattern problems are eliminated.

---

† E. L. Jordan of RCA, using x-ray diffraction, found the surface to consist of elemental boron after a heavy boron tribromide deposition.

[5] H. Rupprecht and G. H. Schwuttke, "Distribution of Boron-Induced Defects in Shallow Diffused Surface Layers of Silicon," *Jour. Appl. Phys.*, Vol. 37, p. 2862, June 1966.

[6] J. Scott, Jr., and J. Olmstead, "A Solid-to-Solid Diffusion Technique," *RCA Review*, Vol. 26, p. 357, Sept. 1965.

[7] R. S. Yatsko and J. S. Kesperio, "A Modified Closed-Box System for the Diffusion of Boron in Silicon," *Jour. Electrochem. Soc.*, Vol. 107, p. 911, 1960.

2. Because each slice of BN serves as a diffusion source for two silicon slices, maximum utilization of the deposition furnace is obtained.

3. The BN source is easily activated by a periodic firing in an oxygen atmosphere. Because negligible material is consumed during any diffusion, the source lasts virtually indefinitely.

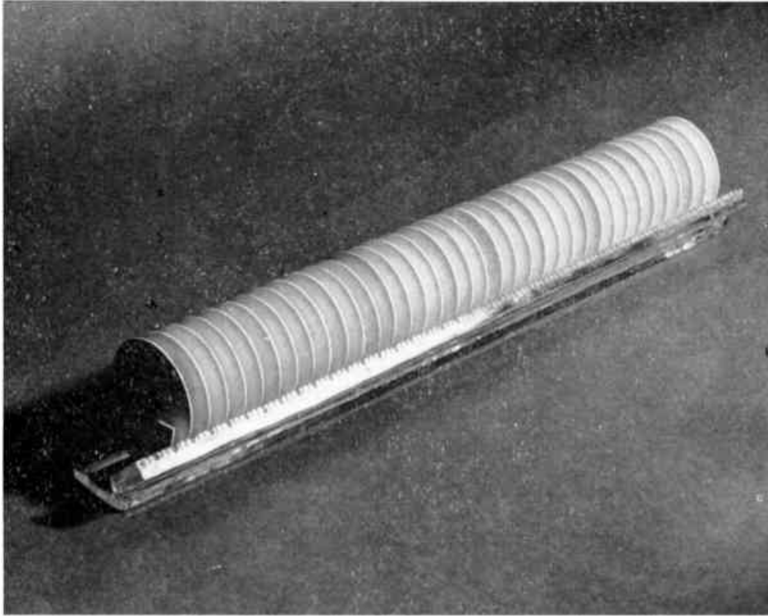


Fig. 1—Boron nitride wafers arranged in quartz boat for oxidation and subsequent use as diffusion sources.

4. The value of F_s during the deposition is controlled solely by the temperature. This value can be kept low to reduce damage to the silicon and the masking oxide.

5. Use of a single inert gas during the deposition eliminates the usual variable of carrier-gas composition.

6. Because the quantity of available boron oxide is limited, the common problem of sticky furnace tubes is eliminated.

7. BN is an inert, easily handled material, available in high purity.

8. The BN system is compatible with other systems and existing equipment.

9. Because the control of low doping levels is excellent, steam is not required during drive-in. As a result, a more meaningful value of sheet resistance after drive-in can be obtained by measurement after deposition.

Experimental Procedure and Results

Boron nitride was obtained as bar stock, 6 inches long and 1.5 inches in diameter, from the Carborundum Corporation. The bar stock was milled into slices 0.025 inch thick with flats on two edges at right angles

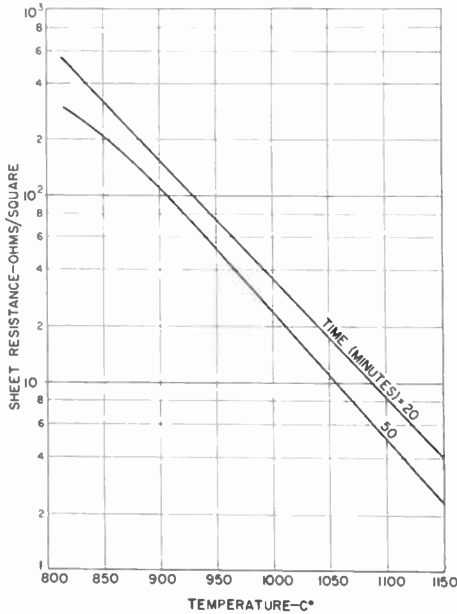


Fig. 2—Deposition sheet resistance as a function of temperature for deposition times of 20 and 50 minutes.

to each other. The quartz boat used to hold the wafers is V-shaped and has eight 0.01-inch-wide slots to the inch, with every third slot enlarged to 0.03 inch. The BN slices are cleaned and then placed in the widened slots, as shown in Figure 1.

Before BN can be used as a diffusion source, the surface must be oxidized. This operation is not critical; heating for an hour at 950°C is sufficient. Periodic reactivation is required, the frequency depending upon the time and temperature at which the wafers are used. If the wafers are in continuous use at high temperature, weekly reoxidation may be required. The need for reactivation is indicated by an increase in the deposition sheet resistance.

The silicon wafers are loaded into the boat, with each wafer facing a BN wafer. The boat is then placed in a furnace for the time and temperature necessary to provide the desired sheet resistance. The curves needed to decide on the deposition schedule were obtained for

temperatures ranging from 815° to 1150°C and times ranging from 10 to 50 minutes. Sheet-resistance data obtained in one set of experiments are shown in Figures 2 and 3. At temperatures of 1100°C and greater, argon was used in place of nitrogen as the ambient gas to prevent the growth of silicon nitride on the 1- to 2-ohm-centimeter n-type silicon wafers. At all temperatures investigated, uniformity from wafer to

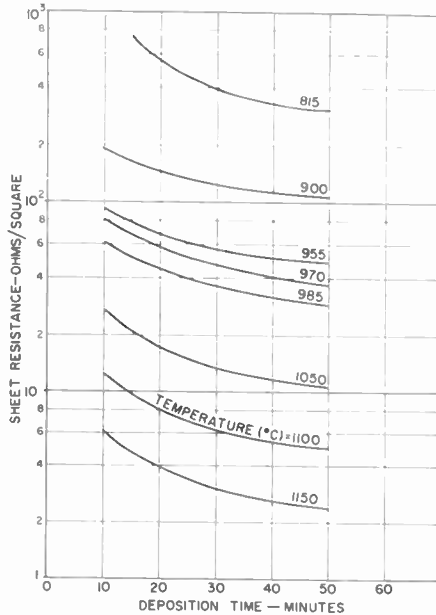


Fig. 3—Deposition sheet resistance as a function of deposition time for several deposition temperatures.

wafer, and across a wafer, was excellent. Variations in sheet resistance were less than ± 5%.

The curves of Figures 2 and 3 show that the boron nitride system is very sensitive to temperature. This sensitivity is not objectionable because modern furnaces are capable of maintaining constant temperature (within 0.5°C) over lengths exceeding 20 inches. If changes are made in furnace design, however, some variation from the curves in Figures 2 and 3 may result because of differences in the recovery time of the furnace (the source temperature is the controlling factor during deposition). An example of the variation that may occur is evident from comparison of Figures 3 and 4. The values shown in Figure 4 were obtained in the initial experimental evaluation of the BN system.

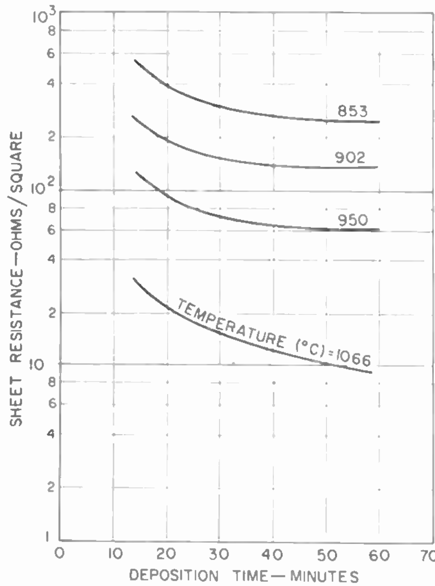


Fig. 4—Deposition sheet resistance as a function of time in an early experimental furnace.

The furnace used was smaller and had a much larger recovery time than the one used in subsequent work.

Acknowledgment

The authors thank M. A. Blumenfeld for permission to use the results shown in Figures 2 and 3, and C. Sickles and S. Muller for their technical assistance.

A PROGRAMMING SYSTEM FOR ELECTRON OPTICAL SIMULATION

By

H. E. Kulsrud*

Summary—A programming system for the RCA 601 computer has been developed and is used to simulate electron optical devices. Computer simulation permits the testing and analysis of designs at a minimal cost and in a short time. The system unifies all computer programs for the simulation and allows their use by the computer novice. Important features of the system are a user-oriented language, simple problem specification, extendability, batching of problems, and speed. Potentials, trajectories, equipotential contours, field lines, Gaussian constants, and aberrations of electrostatic lenses can be calculated.

Introduction

THERE is an ever-increasing need to design electron optical devices to high accuracy. A computer programming system provides a cheap and effective means for designing and analyzing such devices as image tubes, electrostatic lenses, and electron guns for traveling-wave tubes. The device to be analyzed is simulated on the computer and questions about its performance are asked. To utilize the computer as if it were a simple piece of laboratory equipment, the specification of the device and the requests for information about the device should be made in a manner natural to the engineer working on the device. It should not be necessary for the engineer to know very much about computers. The system should be easily extendable so that as new mathematical methods are developed these can be added. Furthermore, self-explanatory error diagnostics should be issued by the system when the user makes a mistake.

A programming system that fits these specifications has been written for the RCA 601 computer. This system for simulating electron-optical devices has many advantages over other specialized problem-solving routines.[1-3] Of particular importance is its use of an electron-

* Formerly with RCA Laboratories, Princeton, New Jersey; now at Yale University, New Haven, Connecticut.

[1] J. E. Boers, "Digital Computer Analysis of Axially Symmetric Electron Guns," *IEEE Trans. Electron Devices*, Vol. ED-12, p. 425, July 1965.

[2] V. Hamza and E. A. Richley, "Numerical Evaluation of Ion-Thruster Optics," NASA Technical Note D-1665, Lewis Research Center, Cleveland, Ohio, May 1963.

[3] Dubois, *et al.*, "Digital Computer Analysis of Axially Symmetric Solid-Beam Electron Guns," RADC-TDR-64-119, University of Michigan, Ann Arbor, Michigan, April 1964.

optics language, generality in the shape of the devices studied, and ease of extendability.

MATHEMATICAL SIMULATION

To study an electron optical device, the shape of the region and the boundary potentials are specified. The program then prepares a computer-memory mapping of the discrete mesh form of the region. Potentials inside the region are found by solving the applicable partial differential equation using relaxation techniques.[4,5] The properties of the potential distribution (and thus of the device) are explored in terms of equipotential contours, field lines, trajectories of electrons with given initial velocities, aberrations, and other characteristics. Output from the system is given in both numeric and graphic form.

A problem is considered to be a set of tasks pertinent to a given device geometry and boundary potentials. However, any computer run of the Electron-Optics System can process many "problems" either related (e.g., a systematic change in a potential value) or unrelated (e.g., an image tube followed by a vidicon). The output of a problem can be saved in case additional tasks are desired in the future.

Tasks are specified by a set of task or function words. On processing a given task word, the system will ascertain whether it is possible to perform the task at the present time (for instance, one cannot trace trajectories unless there is a potential distribution available). Then it will check for any optional features requested and read in the required numerical data for the task. Finally the task will be carried out.

TASKS CARRIED OUT BY THE SYSTEM

The tasks carried out by the system are of four types, as summarized in Table I. Tasks of Type 1 are concerned with creating the mesh mapping in the computer memory. On any first computer run for a device, the mesh must be generated from punched cards via the INPUT CARDS task. Thereafter it is stored on magnetic tape from which it may be obtained on subsequent runs by use of the INPUT TAPE task. The SUBREGION task allows for extracting a subregion of a previously generated map, e.g., in the neighborhood of the cathode.

Type 2 tasks are those that obtain the potential distribution. At the present time, these include the solution of Laplace's equation in

---

[4] D. Young, "Iterative Methods For Solving Partial Difference Equations of Elliptic Type," *Trans. Amer. Math. Soc.*, Vol. 76, p. 92, 1954.

[5] G. E. Forsythe and W. Wasow, *Finite-Difference Methods for Partial Differential Equations*, pp. 242-266, John Wiley and Sons, Inc., New York, 1960.

two dimensions, in three dimensions with cylindrical symmetry, and the solution of Poisson's equation for some special conditions.

Type 3 tasks are those that describe the properties of a given potential distribution. These include the TRAJECTORY task for the

Table I—Tasks Carried Out By The System

Task	Description
TYPE 1: Memory Mapping	
INPUT CARDS	Data read from cards
INPUT TAPE	Previous map read from tape
SUBREGION	Extract a subregion from a map
TYPE 2: Potential Solution	
POTENT	Solve Laplace's equation in cylindrically symmetric region
POTENT XY	Solve Laplace's equation in x-y geometry
POTENT SC	Solve Poisson's equation in cylindrically symmetric region
TYPE 3: Properties of the Potential Solution	
TRAJECTORY	Plot trajectories
CONTOUR	Plot equipotential contours
FIELD LINES	Plot field lines
ABERRATIONS	Plot paraxial rays and calculate aberrations
TYPE 4: System	
REWIND n	Rewind tape n
DUMP n	Write region map on tape n
SKIP m	Skip m maps on the input tape
ADDITION	Go to a user written program
END	Last card of the case

tracing of electron paths, the CONTOUR task which traces equipotential contours, and the FIELD LINES task which traces the field lines. A group of jobs listed under the task word ABERRATIONS produces tracings of the paraxial rays and computes the aberrations of an electrostatic lens.

Type 4 tasks are the executive tasks of the system and provide for rewinding, reading, and writing of magnetic tapes containing the memory map with the internal potentials. The END task allows for the running or batching of several problems on a single computer access. The ADDITION task permits the addition of FORTRAN subroutines coded by the user to a system run.

The electron optics system is run under the RCA 601 FORTRAN operating system and therefore requires no special computer setup.

Specification of the Region

The specification of the region's shape is essentially a geometric problem. It requires that the continuous problem envisaged by the engineer be discretized. The boundary is specified by segments and the end points of these segments are given to the computer. In addition, a code is used to specify whether the potential is constant or varies along the segment and whether the segment is a straight line, arc of a circle, or other conic section. Table II gives boundary data for a typical problem.

Table II—Sample of Line-Segment Specification

r value	z value	Potential	Line code	Comment
252.	0.	.115	0	Cathode line
246.	3.	0.	2	Point
12.	3.	0.	5	Vertical line
11.999	6.	0.	1	
12.	7.	0.	2	A corner
24.	7.	0.	5	
42.	12.4	0.	5	Line at an angle
48.	13.	0.	2	
246.	13.	0.	5	
252.	18.	.014375	0	Gap
252.	48.	.100625	10	Potential varies linearly
252.	54.	.115	0	Electrode
252.	288.	.115	0	
252.	294.	.2340476	0	Gap
252.	498.	24.608762	10	Potential varies linearly
246.	593.	25.115	2	
192.	593.	25.115	5	
188.	498.	25.115	1	
188.	264.	25.115	4	Horizontal line
186.	264.	25.115	0	
126.	264.	25.115	0	
126.	276.	25.115	0	
168.	276.	25.115	0	
168.	756.	25.115	0	Closing the region
0.	756.	25.115	0 1	Last card code

The computer program then generates the discrete representation of the line segment to the requested accuracy. The user specifies the accuracy by first giving the number of mesh lines and the maximum dimensions. An optional feature that the user may request is a graphic plot (via card output) of the input supplied. This provides a means of checking whether the boundary has been properly specified. In the present version a grid of 9000 points or less may be used.

Boundary lines need not fall on the mesh lines of the discrete representation. The program properly adjusts the relaxation formula in the neighborhood of a non-grid boundary. Furthermore, the program examines the range of variables (distance and potential) to carry out all calculations to the required accuracy. Data is usually printed in a fixed-point format rather than floating point format (e.g., .00012 rather than 1.2E-4), as this is more natural to the engineer.

Solution of the Partial Differential Equation

To obtain the potential distribution, it is necessary to solve one of the following three partial differential equations.

Laplace (x-y geometry)

$$\frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial x^2} = 0\,, \tag{1}$$

Laplace (r-z geometry)

$$\frac{\partial^2\phi}{\partial r^2} + \frac{1}{r}\frac{\partial\phi}{\partial r} + \frac{\partial^2\phi}{\partial z^2} = 0\,, \tag{2}$$

Poisson (r-z geometry)

$$\frac{\partial^2\phi}{\partial r^2} + \frac{1}{r}\frac{\partial\phi}{\partial r} + \frac{\partial^2\phi}{\partial z^2} = -\frac{\rho}{\epsilon_0}\,, \tag{3}$$

where ϵ_0 is the dielectric constant. Equation (3) actually involves the solution of Equation (2) many times, since both the potential ϕ and the charge density ρ are unknown. Values for ρ are assumed, the potential is then found, and the values for ρ are corrected. The process continues until the solution for both has converged.[6]

We reduce these partial differential equations. For instance, Equation (3) is reduced to its difference analogue form,

$$4\phi_0 = \phi_1 + \phi_2 + \phi_3 + \phi_4 + \frac{h}{2r_0}(\phi_3 - \phi_4) + \frac{\rho h^2}{\epsilon_0}\,, \tag{4}$$

where ϕ_0 is a point in the region, and ϕ_1, ϕ_2, ϕ_3, and ϕ_4 are its four

[6] P. T. Kirstein and J. S. Hornsby, "An Investigation Into the Use of Iteration Methods for the Analysis of Axially Symmetric and Sheet Beam Electrode Shapes With an Emitting Surface," *IEEE Trans. Electron Devices*, Vol. ED-11, p. 195, May 1964.

equidistant adjacent neighbors, as shown in Figure 1. The mesh size, h, is the distance between points, and r_0 is the r co-ordinate of ϕ_0. Equation (4), or an analogous equation for nonequidistant points and/or axis points, holds for each point on the mesh and, therefore, we obtain a system of linear equations. The linear system is solved

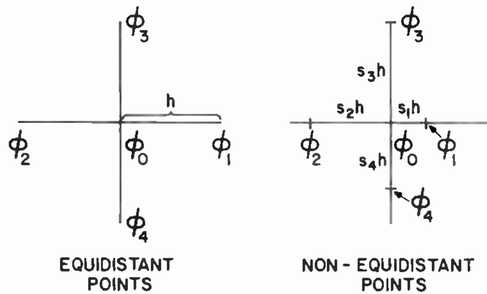


Fig. 1—Graphic representation of difference analogue form of Equation (3).

by the successive over-relaxation method, which is an iterative method. Equation (4) and its three counterparts take the following forms.

Equidistant Off Axis:

$$\phi_0^{(n+1)} = (1-\omega)\,\phi_0^{(n)} + \frac{\omega}{4}\left(\phi_1^{(n)} + \phi_2^{(n+1)} + \left(1+\frac{h}{2r_0}\right)\phi_3^{(n+1)} + \left(1-\frac{h}{2r_0}\right)\phi_4^{(n)} + \frac{\rho_0 h^2}{\epsilon_0}\right). \quad (5)$$

Equidistant On Axis:

$$\phi_0^{(n+1)} = (1-\omega)\,\phi_0^{(n)} + \frac{\omega}{6}\left(\phi_1^{(n)} + \phi_2^{(n+1)} + 4\,\phi_3^{(n+1)} + \frac{\rho_0 h^2}{\epsilon_0}\right). \quad (6)$$

Non-Equidistant Off Axis:

$$\phi_0^{(n+1)} = (1-\omega)\,\phi_0^{(n)} + \frac{\omega}{k_1}\left[\frac{\phi_1^{(n)}}{s_1(s_1+s_2)} + \frac{\phi_2^{(n+1)}}{s_2(s_1+s_2)} + \frac{\left(1+\frac{h}{2r_0}s_4\right)\phi_3^{(n+1)}}{s_3(s_3+s_4)} + \frac{\left(1-\frac{h}{2r_0}s_3\right)\phi_4^{(n)}}{s_4(s_3+s_4)} + \frac{\rho_0 h^2}{\epsilon_0}\right], \quad (7)$$

where

$$k_1 = \left[\frac{1}{s_1 s_2} + \frac{\left(1 + \frac{h}{2r_0}(s_4 - s_3)\right)}{s_3 s_4} \right].$$

Non-Equidistant On Axis:

$$\phi_0^{(n+1)} = (1-\omega)\,\phi_0^{(n)} + \frac{\omega}{k_2}\left(\frac{\phi_1^{(n)}}{s_1(s_1+s_2)} + \frac{\phi_2^{(n+1)}}{s_2(s_1+s_2)} + \frac{2\phi_3^{(n+1)}}{s_3^2} + \frac{\rho_0 h^2}{\epsilon_0}\right). \quad (8)$$

where

$$k_2 = \frac{1}{s_1 s_2} + \frac{2}{s_3^2}.$$

In the above equations, s_1, s_2, s_3, and s_4 are related to h as indicated in the right-hand part of Figure 1, and n is the iteration number. The convergence of the iteration is speeded by proper choice of ω, the optimum relaxation factor.[7]

When the potentials have reached the desired accuracy, usually 10^{-6} times the maximum boundary potential, they are printed. Special features that allow printing of only the axis, no printing, and more or less accuracy can be obtained through commands in the Electron-Optics language. During the solution of Poisson's equation the charge density will also be printed.

PROPERTIES OF THE POTENTIAL SOLUTION

The tasks of Type 3 are formulated as systems of first-order differential equations that can be solved by a standard predictor–corrector method.[8] The equations of motion for trajectories of a given initial velocity are

$$\frac{d^2z}{dt^2} = \frac{e}{m}\frac{\partial\phi}{\partial z}, \quad \frac{d^2r}{dt^2} = \frac{e}{m}\frac{\partial\phi}{\partial r}. \quad (9)$$

The derivatives of ϕ with respect to r and z are found by double linear interpolation within a mesh box. This order of interpolation, which

[7] H. E. Kulsrud, "A Practical Technique for the Determination of the Optimum Relaxation Factor for the Successive Over-Relaxation Method," *Comm. of the ACM*, Vol. 4, No. 4, p. 184, April 1961.

[8] P. Henrici, *Discrete Variable Methods In Ordinary Differential Equations*, pp. 187-202, John Wiley and Sons, Inc., New York, 1962.

requires twelve points, is justified, since the mesh size has been chosen to keep the variation within a box to this accuracy.

The standard mode of operation assumes a flat cathode. Trajectories of a specified velocity, the initial values of dr/dt and dz/dt given, are traced starting at every mesh line along the cathode. Trajectories are terminated when they come within one mesh box of the edge of the region. Special options permit starting trajectories every p^{th} line along the cathode or any place in the region. Curved cathodes may also be used. The end points of trajectories may be saved (on magnetic tape) and later used as starting points for an adjacent region. This feature is useful when the region near the cathode is analyzed in more detail than the region near the anode. The accuracy criterion (four significant figures) can also be increased or decreased.

The output will consist of the position of the electron as it moves through the region, and the derivatives of position with respect to time. The output can be obtained as a printed list or as a graph. A sample of graphic output is given in Figure 2.

Contour tracing is formulated as a pair of first-order differential equations[9] where s is an arc length,

$$\frac{dr}{ds} = \frac{\pm \dfrac{\partial \phi}{\partial z}}{\sqrt{\left(\dfrac{\partial \phi}{\partial r}\right)^2 + \left(\dfrac{\partial \phi}{\partial z}\right)^2}}$$

$$\frac{dz}{ds} = \frac{\mp \dfrac{\partial \phi}{\partial r}}{\sqrt{\left(\dfrac{\partial \phi}{\partial r}\right)^2 + \left(\dfrac{\partial \phi}{\partial z}\right)^2}} . \qquad (10)$$

Similarly the field lines are found from

$$\frac{dr}{ds} = \frac{\mp \dfrac{\partial \phi}{\partial r}}{\sqrt{\left(\dfrac{\partial \phi}{\partial r}\right)^2 + \left(\dfrac{\partial \phi}{\partial z}\right)^2}} \qquad (11)$$

[9] R. W. Klopfenstein, "Zeroes of Non-Linear Functions," *Jour. ACM*, Vol. 8, No. 3, p. 366, 1961.

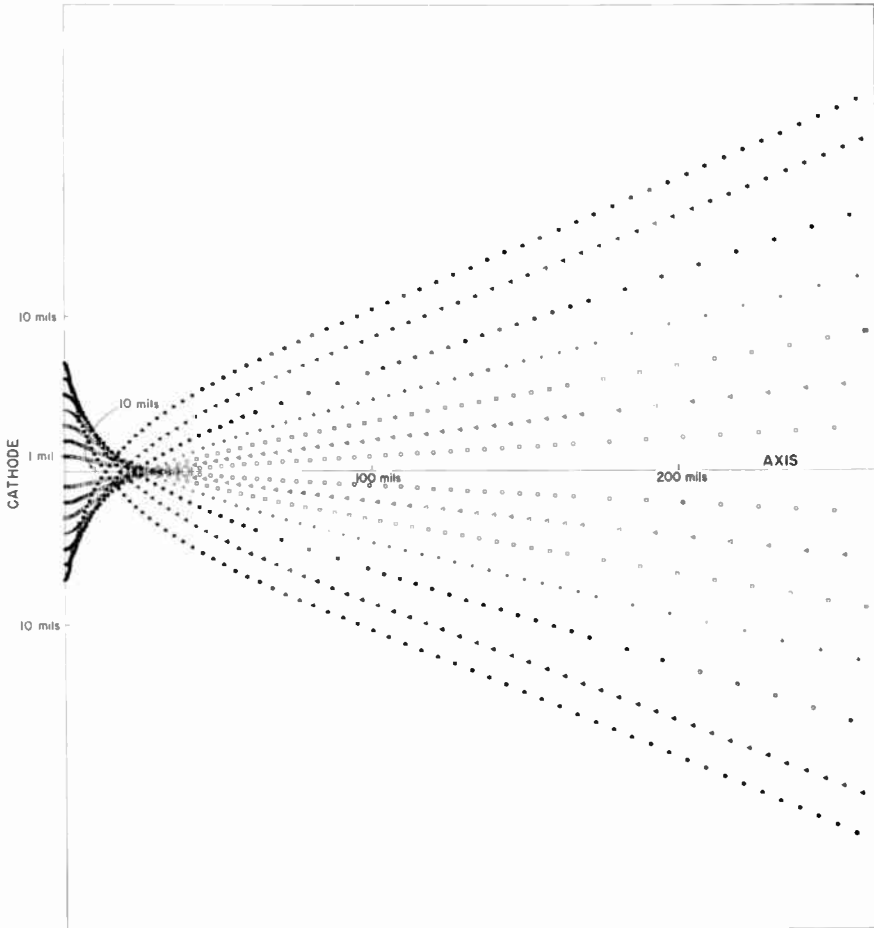


Fig. 2—Sample of graphic output showing plot of trajectories in gun S, 30 × 264 mils.

$$\frac{dz}{ds} = \frac{\mp \dfrac{\partial \phi}{\partial z}}{\sqrt{\left(\dfrac{\partial \phi}{\partial r}\right)^2 + \left(\dfrac{\partial \phi}{\partial z}\right)^2}}.$$

This method is faster than a straight-forward searching technique and the output is in a good form for plotting.

Contours can be traced for a given axial potential value or by specifying where the contour is to start. Field lines are usually begun on the cathode, but can be started at any arbitrary position r, z.

Special Features

The SUBREGION task allows for studying the region of interest accurately when the boundary potentials are known at a large distance. To handle this situation, the use of a coarse mesh over most of the region joined to a fine mesh in the region of interest was suggested by Shortley *et al.*[10] However, the program logic required to attach the edges of the two regions is formidable and time consuming. Some authors have allowed different spacing in the r and the z directions[11] or completely arbitrary spacing for each r and z line.[12] The former solution is not flexible enough whereas the latter is too cumbersome for the inexperienced computer user.

Thus, it appears simpler in concept and faster in time to solve the problem in two or more passes, each pass having the correct size mesh for the amount of accuracy required. For example, first the entire region is run with a coarse mesh, then the region of interest is extracted from the larger region and run with a finer mesh spacing, etc. The computer performs the extraction, and the potentials from the coarse mesh act as a good initial guess for the fine-mesh iteration.

The ADDITION task has proven to be extremely valuable. Besides serving as a means of adding FORTRAN coded programs and even machine language programs to the system, it has functioned as an important debugging device. When new tasks are to be added to the system, they are first tested under the name ADDITION. The Electron Optics System is not changed until a final version of the new task coding has been accepted. Furthermore, suspected errors in the systems can be analyzed and corrected in the same way.

The general parameters of the system, such as the mesh size, number of mesh points, and values of potentials, are available to the user via FORTRAN COMMON storage. These parameters are listed in Table III.

An effort has been made to catch all user errors and print out diagnostic comments. When an error is detected, the "problem" is

[10] Shortley, Darby, Weller, and Gamble, "Numerical Solution of Axisymmetrical Problems with Applications to Electrostatics and Torsion," *Jour. of Appl. Phys.*, Vol. 18, p. 116, Jan. 1947.

[11] R. L. Beurle, W. M. Wreathall, and B. A. Carré, "Research on Electron-Optical Design," English Electric Valve Co. Ltd., England, February 1962.

[12] D. C. Prince, G. Kuskevics, and R. Edwards, "Computer Evaluation of Ion Engines," NAS 8-623, G. E., Cincinnati, Ohio, Nov. 1962.

Table III—COMMON Storage

Name	No. of Words	Use
MODE	1	0 for cylindrical mode, 1 for two dimensional, iteration count for POTENT SC.
NR	1	Number of mesh lines in r direction.
NZ	1	Number of mesh lines in z direction.
NRZ	1	Number of mesh points.
RMAX	1	Maximum actual value of r (in mils, cm, etc.)
ZMAX	1	Maximum actual value of z.
H	1	Mesh size.
RDOT	1	dr/dt at the initial point of a trajectory.
ZDOT	1	dz/dt at the initial point of a trajectory.
FREQ	1	Frequency at which trajectories are to be started.
NQ	1	Length of table of special points.
NPUNCH	1	0 for no punching, 1 for punching cards.
NOALL	1	-1 for no printing of potentials, 0 for printing, 1 for printing the axis only.
FM	3	Format for card punching of trajectories, etc.
NTRACK	1	Point in the REGION table where table of special point distances ends.
POTMAX	1	Maximum value of potential on the boundary.
ALPHA	1	$\omega/4$.
CONVER	1	Number of significant figures in potential solution.
NCURVC	1	0 for flat cathode, 1 for curved cathode.
NADJ	1	0 normal, 1 for ADJ REGION option in ABERRATIONS task. -1 for SAVE END option, -2 for ADJ REGION and SAVE END.
NAME	10	80 characters of problem identification.
REGION	9000	Region mesh picture of length NRZ, table of special point distances, table of special point potentials, space charge table if POTENT SC.
MTAPE	1	Tape number for Optics Commands input data.
NTAPE	1	Tape number for INPUT CARDS and SUBREGION data.

Table IV—Possible Diagnostic Comments

Illegal first control word
Repeated boundary point
Program logic error
Input data not valid
Ltype not programmed
Too many mesh points
Too many special points
R-Z scaling too large
Bad input word
Values too large to be printed
Unclosed region at $r = -$
Boundary not on coarse mesh line
Machine or program error in spherical aberration
Divergence during relaxation
Logic error in subregion program
Terminated by operator

abandoned and the END card searched for. Processing continues with the next card, new "problem", after the END. Certain errors, however, are so extreme that they will completely terminate a machine run (e.g., errors that may destroy the system itself.) Table IV gives a listing of the possible diagnostic comments.

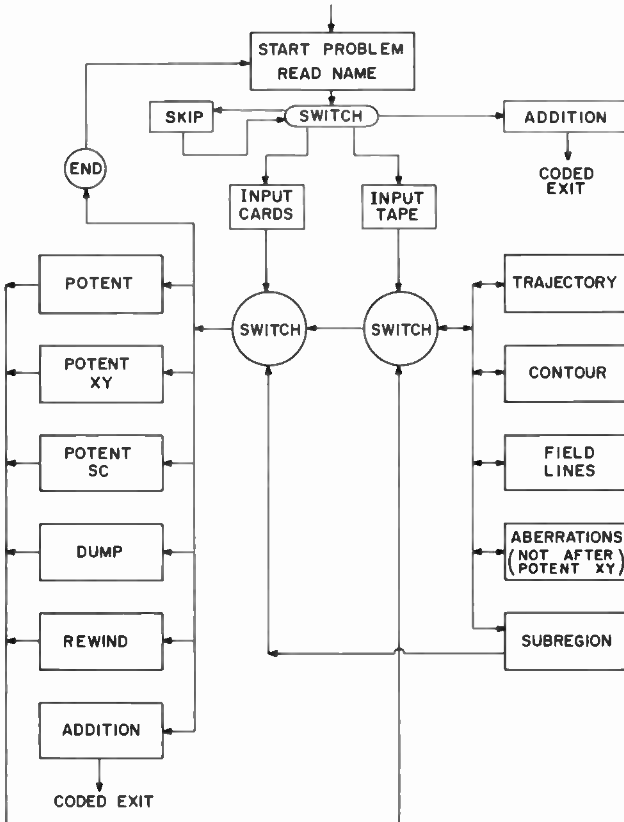


Fig. 3—General flow chart.

A listing of the tasks and special options currently included in the system is given in Table V. A general flow chart, Figure 3, gives a pictorial representation of the tasks and when they may be used.

Accuracy and Timing

As yet there has been no complete mathematical theory of the size of mesh required to produce a solution of Laplace's equation to a given

accuracy. The user must chose the mesh size on the basis of experience with his type of problem. If no knowledge is available a problem should be run twice, the second time with half the mesh size, and the results compared. Unless instructed otherwise, the Electron Optics

Table V—Tasks and Special Options Currently Included in System

Task Code Words	Options	Numeric Data
INPUT CARDS	PUNCH	No. of R lines, no. of Z lines, R_{max}; line segments.
INPUT TAPE		
DUMP *n*		
REWIND *n*		
SKIP *n*		
POTENT POTENT XY POTENT SC	PRINT AXIS NO PRINT CUT OFF ITERATIONS *n* SIG FIGURE *n*	$T_c k/c$, field zero point (applies only to POTENT SC)
TRAJECTORY	CURVED CATH MAX TRAJECT *n* START TRAJ *p* SAVE END READ CARDS ADJ REGION	$\dot{r}$, $\dot{z}$, or $r, z, \dot{r}, \dot{z}$ z edge
CONTOUR	CURVED CATH START CONTOUR *p* READ CARDS	Potential or r, z
FIELD LINES	CURVED CATH START LINE *p* READ CARDS	r, z
ABERRATIONS	PUNCH SAVE END READ CARDS ADJ REGION	 z of cathode z edge
SUBREGION	PUNCH READ CARDS	r_{min}, z_{min}, z_{max}, increase in line number.
ADDITION		
END		

System will find potentials to six significant figures and carry out the tasks of Type 3 to four significant figures.

The normal problem handled by the system runs between one and ten minutes on the RCA 601 computer. A typical problem with 4000 points and plotting twenty trajectories requires two minutes of computer time. Ten minutes is used by a problem with 9000 points, a

subregion extraction and solution, and processing of trajectories and contours through both regions. Table VI gives the input data for a sample problem for finding potentials and tracing equipotential con-

Table VI—Sample Problem: Post Accelerating Anode Potential Field as Given in Figure 4. Find the Potentials in the Region and Plot the Equipotential Lines of .95, .9, .8, .6, .5, .4, .3, .2, .1, .05 and .03 volts.

Card					Comment
*JOB SAMPLE 0008t050					FORTRAN System-time and pages.
*PAUSE MOUNT OPTICS SYSTEM ON RPD 1					Command to machine operator.
*EXECUTE					FORTRAN System command.
*COMPILE					FORTRAN System command.
CALL CHAIN (1,1)					FORTRAN language to call Optics.
END					FORTRAN language end.
*DATA					FORTRAN System data command.
SAMPLE PROBLEM, POST ACCELERATING ANODE POTENTIAL FIELD					
INPUT CARDS					Optics system task word.
35.	41.	17.			No. of r lines, z lines, $r_{max.}$ value.
1.	0.	0.	0		First region shape card.
1.	4.	0.	0		
17.	17.	1.	10		Linear potential variation.
17.	20.	1.	0		
0.	20.	1.	0	1	Last region card.
POTENT					Compute Potentials.
CONTOUR					Equipotentials.
.95					Potential values to be traced.
.90					
.80					
.60					
.50					
.40					
.30					
.20					
.10					
.05					
.03				1	Last Potential card.
END					End of problem
*END FILE					End of Optics System and Fortran job.

tours. It took less than a minute on the computer. The graphic solution of this problem is shown in Figure 4.

Conclusion

The system described in this paper reflects the current status of computer simulation of electron optical devices, but allows for the in-

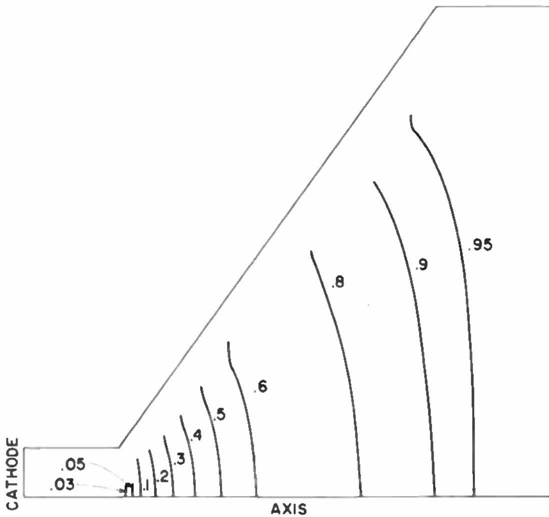


Fig. 4—Equipotential lines of post accelerating-anode potential field.

corporation of new methods as they become available. Older methods are retained in case regression is necessary. Accessing the system through a user-oriented language permits use by the computer novice and the engineer unfamiliar with the details of the mathematical techniques. Enlarging the number of users of the system reveals new applications for the system. The system developers receive new ideas on how to improve and extend the system. Thus the system approach serves not only to consolidate current knowledge but to point to new areas of research both on mathematical methods and devices.

ACKNOWLEDGMENT

The author is indebted to T. M. Stiller for his many useful suggestions on 601 techniques and for adding some special programs to 601 FORTRAN to facilitate writing of the Electron Optics System.

A COLLISION-INDUCED INSTABILITY IN SEMICONDUCTOR PLASMAS*

By

BRUCE B. ROBINSON

RCA Laboratories
Princeton, N. J.

Summary—The stability of a system consisting of electrons and holes with relative drift transverse to a static magnetic field is considered. Collision-free theories have predicted unstable interaction between the electron and hole hybrid space-charge waves. It is noted that the addition of collisions not only modifies the hybrid–hybrid interaction but also induces an entirely new mode of instability. An approximate analytic solution that reveals the nature of this collision-induced mode is presented. Since both modes may be pertinent to the theory of microwave emission from InSb, exact numerical solutions including collisions and diffusion are presented.

INTRODUCTION

IT HAS BEEN KNOWN for some time that the addition of dissipation into a hydrodynamic system not only modifies instabilities that exist without dissipation but, under some conditions, actually induces new instabilities.[1] Since the initial paper by Furth, Killeen, and Rosenbluth,[2] considerable analysis has been made of finite-resistivity instabilities in hydromagnetics. Once again, in addition to modifying zero-resistance instabilities, finite resistance induces entirely new modes of instability. In these cases, the introduction of collisions permits more general sets of motions by which the system can reach lower energy states. For example, in the hydromagnetic case the fluid is no longer "frozen" to the field lines when collisions are present.

In the recent literature concerning streaming instabilities in solids, a number of so-called "resistive" instabilities have been discussed. It has been noted by Vural and Bloom[3] that some of these instabilities are really collision-modified and not collision-induced. They also noted

* The research reported in this paper was sponsored by the Air Force Systems Command, Rome Air Development Center, Griffis Air Force Base, New York.

[1] C. C. Lin, *The Theory of Hydrodynamic Stability*, Cambridge University Press, 1955.

[2] H. P. Furth, J. Killeen, and M. N. Rosenbluth, "Finite-Resistivity Instabilities of a Sheet Pinch," *Phys. of Fluids*, Vol. 6, p. 459, April 1963.

[3] B. Vural and S. Bloom, "Streaming Instabilities of Solids and the Role of Collisions," *IEEE Trans. on Electron Devices*, Vol. ED-13, p. 57, Jan. 1966.

the possibility of a collision-induced mode. Shimomura *et al*[4] and others have discussed the possible role of collisions in inducing instability by creating a population inversion in the electron velocity distribution. The purpose of this paper is to describe a collision-induced instability that occurs in the analysis of an infinite system consisting of two streams of charged particles in the presence of a static magnetic field perpendicular to the streaming direction. This work is based on the two-fluid model of an electron–hole plasma and is, therefore, independent of the detailed structure of the velocity distributions. As in the hydrodynamic and hydromagnetic cases, collisions induce a more general set of motions.

Dispersion Relation

Consider an infinite system consisting of electrons and holes streaming along the z axis in the presence of a static magnetic field along the x axis. If one linearizes the two-fluid model of the electron–hole plasma, assumes perturbed quantities to vary as $\exp(i\omega t - ikz)$ and makes the quasi-static approximation, one arrives at the dispersion relation

$$0 = 1 + \sum_a \frac{\omega_{pa}^2 \Omega_a}{(\omega - ku_a)\,[\omega_{ca}^2 - \Omega_a^2] + k^2 \Omega_a \dfrac{V_{\theta a}^2}{2}} \tag{1}$$

where

$$\Omega_a \equiv \omega - ku_a - i\nu_a \,.$$

The summation is over electrons and holes, $\omega_{pa} \equiv [n_a e^2/(\epsilon_L m_a)]^{1/2}$ is the plasma frequency of carrier a, u_a is the drift velocity, ν_a is the collision frequency, $V_{\theta a} \equiv (2kT_a/m_a)^{1/2}$ is the thermal velocity, ω_{ca} is the cyclotron frequency, and ϵ_L is the dielectric constant of the lattice. Hasegawa[5] has discussed this problem making a number of approximations to make it analytically tractable. It is easily shown that his "resistive" instability is a collision-modified version of the hybrid–hybrid mode that occurs without collisions. Under suitable conditions, the dispersion relation also admits of a collision-induced mode, the nature of which can be revealed with approximate analytic solutions.

---

[4] N. Shimomura, K. Mitani, and S. Tanaka, "Collision-Induced Instability in a Magnetoactive Plasma," *Jour. Phys. Soc. Japan*, Vol. 21, p. 1372, 1966.

[5] A. Hasegawa, "Microinstabilities in Transversely Magnetized Semiconductor Plasmas," *Jour. Appl. Phys.*, Vol. 36, p. 3590, Nov. 1965.

If one assumes no collisions ($\nu_+ = \nu_- = 0$), no diffusion ($V_{\theta+} = V_{\theta-} = 0$), and no hole drift ($u_+ = 0$), the dispersion relation is cubic in k and yields the solutions

$$ku = \omega \pm \sqrt{\omega_{c-}^2 + \omega_{p-}^2 (\omega_{c+}^2 - \omega^2)/(\omega_{c+}^2 - \omega^2 + \omega_{p+}^2)} \qquad (2)$$

and

$$ku = \omega . \qquad (3)$$

Solution (3) is invariably ignored in the literature but plays an important role when collisions are present. One of the above solutions is unstable when

$$\omega_{c+}^2 + \frac{\omega_{p+}^2 \omega_{c-}^2}{(\omega_{c-}^2 + \omega_{p-}^2)} < \omega^2 < \omega_{p+}^2 + \omega_{c+}^2 . \qquad (4)$$

When electron collisions are introduced, Equation (1) is still cubic but now has two potentially unstable modes. If one assumes that ν_- is smaller than the other frequencies involved and expands about the zero-collision solutions, the first two solutions (2) are only quantitatively modified; however, solution (3) can now become unstable. If one assumes $\omega - ku \simeq 0$, and $\nu_- << \omega_{c-}$, Equation (1) becomes

$$0 = 1 + \frac{\omega_{p-}^2 (\omega - ku_- - i\nu_-)}{(\omega - ku_-)\, \omega_{c-}^2} + \frac{\omega_{p+}^2}{\omega_{c+}^2 - \omega^2}, \qquad (5)$$

from which one obtains

$$ku_- = \omega - \frac{i\nu_- \omega_{p-}^2}{\omega_{c-}^2 \left[1 + \dfrac{\omega_{p+}^2}{\omega_{c+}^2 - \omega^2} + \dfrac{\omega_{p-}^2}{\omega_{c-}^2} \right]} . \qquad (6)$$

Equation (6) indicates a convective instability when

$$\nu_- \neq 0$$

and

$$\omega^2 < \omega_{c+}^2 + \frac{\omega_{p+}^2 \omega_{c-}^2}{\omega_{c-}^2 + \omega_{p-}^2} . \qquad (7)$$

With a little more algebra, one may include hole collisions and, assum-

ing $\omega_{c+} << \nu_+$, obtain the conditions

$$\nu_- \neq 0$$

and

$$\omega^2 < \frac{\omega_{p+}{}^2\,\omega_{c-}{}^2}{\omega_{c-}{}^2 + \omega_{p-}{}^2} - \nu_+{}^2 \qquad (8)$$

for instability.

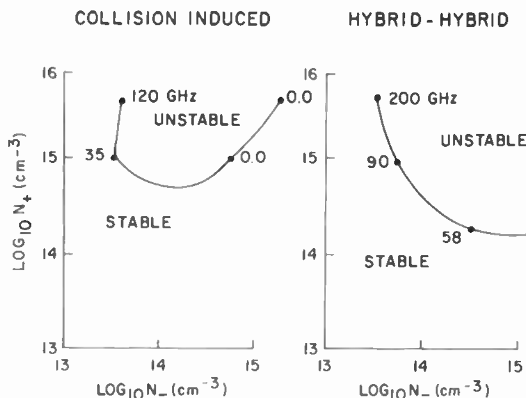


Fig. 1—Onset conditions, $B = 1$ kilogauss, $\nu_+ = \nu_- = 3 \times 10^{11}$ sec^{-1}, the effective mass $m_-^* = 0.014$ of the electron mass m_o, $m_+^* = 0.4\ m_o$, $T_+ = T_- = 260°$ K, $u_- = V_{\theta-}$.

The complete dispersion relation, Equation (1), has been solved numerically without approximation using parameters appropriate for InSb.[6] Both the hybrid–hybrid mode and the collision-induced mode have been studied including collision and diffusion effects. The resulting onset conditions for a typical set of parameters are shown in Figure 1, with several onset frequencies indicated.

Discussion

In order to understand this collision-induced instability, it is useful to consider the problem from the mode interaction point of view.[7] In this picture, the hole and electron fluids are considered as semi-

[6] G. A. Swartz and B. B. Robinson, "Coherent Microwave Emission from Indium Antimonide Structures," *Appl. Phys. Letters*, Vol. 9, p. 232, Sept. 1966.

[7] B. Vural and M. C. Steele, "Possible Two-Stream Instabilities of Drifted Electron-Hole Plasmas in Longitudinal Magnetic Fields," *Phys. Rev.*, Vol. 139, p. A300, July 1965.

independent systems that are weakly coupled. Each fluid supports a set of plasma waves. When their relative drift is sufficient to Doppler shift an electron wave into synchronism with a hole wave, an unstable resonant coupling between the two systems is possible under appropriate conditions determined by energy considerations. When no collisions are present, each fluid supports waves described by the relation

$$0 = 1 + \frac{\omega_{pa}^2 (\omega - ku_a)}{(\omega - ku_a) [\omega_{ca}^2 - (\omega - ku_a - i\nu_a)^2]} \tag{9}$$

where diffusion has been omitted. This relation yields three solutions;

$$ku = \omega \pm \sqrt{\omega_{ca}^2 + \omega_{pa}^2} \tag{10}$$

and

$$ku = \omega \, . \tag{11}$$

The first two solutions are the familiar fast and slow hybrid waves. The usual zero-collision two-stream instability shown in Equation (2) is due to the unstable interaction between the fast hybrid mode of the holes and the slow hybrid mode of the electrons. Since the third solution (11) is not discussed in the literature, a few words about its nature are in order.

Consider the electron fluid as seen in its rest frame with respect to drift along the z axis. There are two steady-state solutions to the fluid equations;

$$0 = \frac{e}{m}\mathbf{E} + \frac{e}{m}\mathbf{v} \times \mathbf{B} \tag{12}$$

$$\nabla \cdot (n\mathbf{v}) = 0 \tag{13}$$

$$\nabla \cdot \mathbf{E} = -\frac{ne}{\epsilon}, \tag{14}$$

of the form e^{-ikz}. In addition to the usual homogeneous-density stationary solution, one may obtain

$$\mathbf{E} = E_z \, e^{-ikz} \, \hat{e}_z \tag{15}$$

$$v_y = \frac{E_z}{B} e^{-ikz} \tag{16}$$

$$n = i \frac{\epsilon}{e} k E_z e^{-ikz} . \quad (17)$$

That is to say, an inhomogeneous-space-charge steady-state solution exists. This is possible because the restoring electric fields can be balanced by a Lorentz force created by a spatially varying transverse motion of the electron fluid. If this steady state is drifted along the z axis, it takes the form of a wave traveling at the electron drift velocity and described by the dispersion relation (11). The a-c currents associated with this wave are transverse to all of the electric fields associated with the coupling problem that are parallel to the direction of propagation. As the dispersion relation (3) reveals, this wave does not interact at all with the hole system. When collisions are introduced, the steady-state solution becomes

$$\mathbf{E} = E_z e^{-ikz} \hat{e}_z \quad (18)$$

$$n = i \frac{\epsilon}{e} k E_z e^{-ikz} \quad (19)$$

$$v_x = 0 \quad (20)$$

$$v_y = \frac{e}{m} \frac{\omega_{c-}}{\omega_{c-}^2 + \nu_-^2} E_z e^{-ikz} \quad (21)$$

and

$$v_z = -\frac{e}{m} \frac{\nu_-}{\omega_{c-}^2 + \nu_-^2} E_z e^{-ikz} . \quad (22)$$

The collisions have not only modified the y component of a-c current but have also introduced a component along the direction of propagation that can interact with the electric fields. As the dispersion relation (6) demonstrates, an unstable interaction with the fast hole-hybrid wave is now possible.

Conclusion

The introduction of collisions into an electron–hole system with relative drift transverse to a static magnetic field modifies the usual hybrid–hybrid instability. In addition, the collisions induce an entirely new mode of instability. The role of collisions in this new mode is

essentially catalytic. They do not create new lower energy states, but merely modify the possible motions of the system, opening previously unavailable paths to existing lower energy states. Calculations using realistic parameters for InSb indicate that both the collision-modified and the collision-induced modes should be observable. In particular, frequencies predicted for the collision-induced mode agree with those of the coherent microwave radiation obtained from 5×10^{14} cm^{-3} p-InSb (20-33GHz).

RCA Technical Papers

First Quarter, 1967

Any request for copies of papers listed herein should be addressed to the publication to which credited.

"Analysis of the Negative Resistance Associated with the Semiconductor-Insulator-Semiconductor (SIS) Structure," J. Shewchun, *Jour. Appl. Phys.* (January) 1967

"Far-Field Pattern of Electron-Bombarded Semiconductor Lasers," F. H. Nicoll, *Proc. IEEE* (Letters) (January) 1967

"GaAs Thin-Film Solar Cells," P. Vohl, D. M. Perkins, S. G. Ellis, R. R. Addiss, W. Hui, and G. Noel, *Trans. IEEE PTGED* (January) 1967

"Instrumentation for the Automatic Measurement of MIS Capacitance vs. Bias Characteristics," K. H. Zaininger, *SCP and Solid State Technology* (January) 1967

"A Light-Activated Semiconductor Switch," D. Meyerhofer, A. S. Keizer, and H. Nelson, *Jour. Appl. Phys.* (January) 1967

"Multiple-Idler Parametric Amplifiers," R. L. Ernst, *Trans. IEEE PTGMTT* (January) 1967

"Thermal Regeneration in Power-Dissipating Elements," C. F. Wheatley, *The Electronic Engineer* (January) 1967

"The V-2 Panel as a Side-Mounted Antenna," R. N. Clark and A. L. Davidson, *Trans. IEEE PTGBC* (January) 1967

"Afterpulses in Photomultipliers," G. A. Morton, H. M. Smith, and R. Wasserman, *Trans. IEEE PTGNS* (February) 1967

"BLIP Condition in Point-to-Point Optical Communication," H. S. Sommers, Jr., and E. K. Gatchell, *Proc. IEEE* (February) .. 1967

"Circuit Aspects of Transistor Parametric Frequency Doublers," A. P. Anderson, *Trans. IEEE PTGED* (February) 1967

"Construction of Long Life Argon Lasers," K. G. Hernqvist and J. R. Fendley, Jr., *IEEE Jour. Quantum Electronics* (February) .. 1967

"Detection of Selenium Clustering in GaAs by Transmission Electron Microscopy," M. S. Abrahams, C. J. Buiocchi, and J. J. Tietjen, *Jour. Appl. Phys.* (February) 1967

"Effect of 'a' Domains on the Movement of 'c' Domain Walls in $BaTiO_3$," D. R. Callaby, *Jour. Appl. Phys.* (February) 1967

"The Effect of Various Design Parameters on the Performance of Medium-Power Traveling-Wave Tubes," M. J. Schindler, *Trans. IEEE PTGED* (February) 1967

"Electrical Properties of $B\text{-}Ag_2Te$ and $B\text{-}Ag_2Se$ from 4.2° to 300°K," R. Dalven and R. Gill, *Jour. Appl. Phys.* (February) 1967

"An Extension of Hazony and Nain's Synthesis Procedure for n-Ports," A. A. Clark, *Proc. IEEE* (Letters) (February) 1967

"Extraneous Light Emission from Photomultipliers," H. R. Krall, *Trans. IEEE PTGNS* (February) 1967

"Failure-Correction Decoding," H. E. White, *Trans. IEEE PTGCOM* (February) 1967

"Handy Device for Crystal Cutting," M. Simhony, *Rev. Sci. Instr.* (Notes) (February) 1967

"Isolation of Lossy Transmission Line Hybrid Circuits," R. M. Kurzrok, *Trans. IEEE PTGMTT* (Correspondence) (February) .. 1967

"A Note on State Minimization of Asynchronous Sequential Functions," M. C. Paull and Coauthor, *Trans. IEEE PTGEC* (February) 1967

"Oxygen as a Beneficial Additive in Cesium Thermionic Energy Converters," J. D. Levine and F. E. Gelhaus, *Jour. Appl. Phys.* (Communications) (February) 1967

"Photomultipliers for Scintillation Counting in Severe Environments," F. A. Helvy and R. M. Matheson, *Trans. IEEE PTGNS* (February) 1967

"Signal-To-Noise Measurements in Audio Tape Recorders," *Broadcast News* (February) 1967

"Some Aspects of Binaural Sound," C. J. Hirsch, *IEEE Spectrum* (February) 1967

"Instability of Electron-Hole Plasma Pinch in Longitudinal Magnetic Fields," K. Ando and M. Glicksman, *Phys. Rev.* (10 February) 1967

"Isopin Formulation of the Theory of a Granular Superconductor," R. H. Parmenter, *Phys. Rev.* (10 February) 1967

"Sensitive Tunable Acoustical Phonon Detector," C. H. Anderson and E. S. Sabisky, *Phys. Rev. Letters* (13 February) 1967

"Measurements of Electron Impact Excitation Cross Sections of Laser States of Argon (II)," J. M. Hammer and C. P. Wen, *Jour. Chem. Phys.* (15 February) 1967

"Ultrasonic Measurements in Single-Crystal Nb_3Sn," K. R. Keller and J. J. Hanak, *Phys. Rev.* (15 February) 1967

"Measurement of $2e/h$ Using the AC Josephson Effect and Its Implications for Quantum Electrodynamics," B. N. Taylor and Coauthors, *Phys. Rev. Letters* (20 February) 1967

"$CaF_2:Dy^{2+}$ Lassers," R. J. Pressley and J. P. Wittke, *IEEE Jour. Quantum Electronics* (March) 1967

"Computer-Communications: Its Impact on Military Applications," W. R. Lonergan, *Signal* (March) 1967

"A Crossed-Field Gain Mechanism Employing Discontinuities," B. Hershenov, *Trans. IEEE PTGED* (Correspondence) (March) 1967

"The Deposition of Vitreous Silicon Dioxide Films from Silane," N. Goldsmith and W. Kern, *RCA Review* (March) 1967

"Design Trade-Offs for r-f Transistor Power Amplifiers," R. Minton, *EE* (March) 1967

"Determinants of Electronically Steerable Antenna Arrays," W. T. Patton, *RCA Review* (March) 1967

"Experimental Investigation of Multiwatt Argon Lasers," I. Gorog and F. W. Spong, *RCA Review* (March) 1967

"Holography: Recording Light in Three Dimensions," B. Shore, *Signal* (March) 1967

"Comments on 'A Large Signal Analysis Leading to Intermodulation Distortion Prediction in Abrupt Junction Varactor Upconverters,'" S. M. Perlow and B. S. Perlman, *Trans. IEEE PTGMTT* (Correspondence) (March) 1967

"Magnetic Field Effect on Acoustoelectric Gain in Semiconductors," M. C. Steele, *RCA Review* (March) 1967

"Microwave Power Coupling to the Helicon Mode in Indium Antimonide," G. A. Swartz, *RCA Review* (March) 1967

"New Tools Help Preserve Connector Reliability," B. Greenspan, *EE* (March) 1967

"A New Type of Latching Switchable Ferrite Junction Circulator," W. W. Siekanowicz and W. A. Schilling, *Proc. IEEE* (Letters) (March) 1967

"Printed English Compression by Dictionary Encoding," H. E. White, *Proc. IEEE* (March) 1967

"Sources in Plasma," M. P. Bachynski, *RCA Review* (March) 1967

"UV-Sensitive Photoemitters," A. H. Sommer, *RCA Review* (March) 1967

"Computer Simulation of the Magnetic Dipole Interaction Problem," R. Moskowitz and Coauthor, *Jour. Appl. Phys.* (1 March) 1967

"Ferromagnetic Resonance of Single Crystals of $CdCr_2S_4$ and $CdCr_2Se_4$," S. B. Berger and H. L. Pinch, *Jour. Appl. Phys.* (1 March) 1967

"Magnetic Properties of the Systems $H_gCr_2S_4$-$CdCr_2S_4$ and $ZnCr_2Se_4$-$CdCr_2Se_4$," P. K. Baltzer, M. Robbins, and P. J. Wojtowicz, *Jour. Appl. Phys.* (1 March) 1967

"Semiconducting and Optical Properties of Ferromagnetic $CdCr_2S_4$ and $CdCr_2Se_4$," H. W. Lehmann and G. Harbeke, *Jour. Appl. Phys.* (1 March) 1967

"High-Temperature Susceptibility of Heisenberg Ferrimagnets," P. J. Wojtowicz, *Phys. Rev.* (10 March) 1967

"Surface Impedance in the Surface Superconducting State," A. Rothwarf, J. I. Gittleman, and B. Rosenblum, *Phys. Rev.* (10 March) 1967

"Band Structure of Spinel-Type Semiconductors," W. Rehwald, *Phys. Rev.* (15 March) 1967

"Damping of Waves in Electron Beams," M. Caulton, *Jour. Appl. Phys.* (15 March) 1967

"Electrical Properties of a Degenerate Intrinsic Semiconductor," R. Dalven, *Jour. Appl. Phys.* (15 March) 1967

"Electrical Properties of Silicon Films Grown Epitaxially on Sapphire," D. J. Dumin, *Jour. Appl. Phys.* (15 March) 1967

"Oscillatory Modes Associated with One Carrier Transient Space-Charge-Limited Currents," R. B. Schilling and Coauthor, *Jour. Appl. Phys.* (15 March) 1967

"Photoemission of Holes from Metals into Anthracene," R. Williams and J. Dresner, *Jour. Chem. Phys.* (15 March) 1967

"Preparation of [100]-Oriented Foils of GaAs for Transmission Electron Microscopy," C. J. Buiocchi, *Jour. Appl. Phys.* (Communications) (15 March) 1967

"Extending IC Technology to Microwave Equipment," H. Sobol, *Electronics* (March 20) 1967

"Getting the Most Out of Circuits with Dielectric Isolation," A. I. Stoller, J. A. Amick, and N. E. Wolff, *Electronics* (March 20) 1967

"Direct Measurement of Hot Electron-Phonon Interactions in GaP," B. F. Williams and R. E. Simon, *Phys. Rev. Letters* (27 March) 1967

"Space-Charge-Limited Current Transient Including Diffusion," R. B. Schilling and Coauthor, *Phys. Rev. Letters* (27 March) 1967

AUTHORS

H. AMEMIYA obtained his BSEE degree from the University of Tokyo in 1948 and a MS degree from Cornell University in 1952. He joined RCA in 1957 and has been engaged in a number of projects, including developmental work on a remote-control system for the gas industry, and the design of a magnetic-drum memory system for the RCA 110 Industrial Computer. He is currently engaged in advanced memory projects at RCA Defense Electronic Products.

MARTIN CAULTON received the Bachelor's degree in 1950, the M.S. in 1952, and the Ph.D. degree in 1954, all in Physics, from Rensselaer Polytechnic Institute. He was an instructor in Physics from 1950 to 1953 while at Rensselaer, and did part of his doctoral work at the Brookhaven National Laboratories, in high-energy nuclear physics, where he was research associate from 1953 to 1954. From 1954 to 1955 he was a Fullbright scholar at the Imperial College of Science and Technology in London. From 1955 to 1958 he was a member of the Technical Staff at Bell Telephone Laboratories working in research and development on low-noise microwave tubes. In 1958 he became assistant professor of physics at Union College, Schenectady, New York. Since joining RCA Laboratories in 1960, he has been engaged in work on microwave power tubes, multivelocity flow problems in electron beams, and microwave solid-state devices. Dr. Caulton is co-author of a textbook on physical electronics and is also adjunct professor of electrical engineering at Drexel Institute of Technology.

NORMAN GOLDSMITH received the B.A. degree with honors in chemistry form Hunter College in 1959 and the M.S. degree in physical chemistry from Stevens Institute of Technology in 1964. He joined the Electronic Components and Devices Division of RCA in 1959 working on elemental purification and the growth of bulk and epitaxial gallium arsenide. More recently he has been active in all phases of semiconductor materials and process technology including diffusion, epitaxial growth, vapor deposition, surface stabilization, and development of nondestructive measurement techniques. Mr. Goldsmith is a member of The American Chemical Society and The Electrochemical Society.

A. G. HOLMES-SIEDLE is a B.A. of Trinity College, Dublin (1954), and a Ph.D. of Cambridge University (1958). He conducted post-doctoral research at the Cambridge University Chemical Laboratories from 1958 to 1960, studying the transfer of energy within biological and chemical systems. From 1960 to 1962, he worked at Hawker-Siddelay Aviation on communication satellite systems and lunar vehicles. Dr. Holmes-Siedle joined RCA in 1962, and has been involved in studying both the fundamental and practical problems of radiation environments in regard to basic damage processes in materials, modification of materials for radiation hardness, the effect of the space radiation on space systems and components, and studies of the effects of intense neutron-gamma pulses on semiconductor devices. He is leader of the Radiation Physics group of the Astro-Electronics Division. Dr. Holmes-Siedle is a Fellow of the British Interplanetary Society, a member of the Institute of Electrical and Electronics Engineers, and is listed in American Men of Science.

KAARE KARSTAD is a graduate from Norway's Institute of Technology, Trondheim 1942. As a postgraduate student he worked in 1943 as an instructor in radio engineering at the university. In early 1944, he was employed in the area of remote control at the ASEA concern in Vesteraas, Sweden. After wartime experience in the technical branch of under-water sound, he was from 1946 until 1952 in charge of the Sonar Section of the Royal Norwegian Navy. From 1948 to 1950, on leave of absence and as an Honorary Fellow of the American Scandinavian Foundation, Mr. Karstad worked at the RCA Laboratories, Princeton, N. J. on television problems. He joined RCA Laboratories in 1952, doing research in the field of color television. He has worked in the field of electronic countermeasures techniques, in the area of infrared systems, and has done work on instrumentation related to research in magnetics and thin films. More recently, he has studied memory circuits and devices, and the application of ultrasonics in the memory field. He is a Senior Member of Institute of Electrical and Electronics Engineers and a member of Sigma Xi.

FAYSEL S. KEBLAWI received the BSEE degree in February 1957 from the American University of Beirut, Lebanon. He received the MSEE degree from North Carolina State University in 1963, and the Ph.D. degree from the same university in 1965. Dr. Keblawi joined RCA Astro-Electronics Division in 1966, and has been working on deep-space communication techniques and associated problems, with primary emphasis on the Voyager Spacecraft and probe Lander missions. Before joining RCA, Dr. Keblawi was engaged in consulting work for Kellog, the transmission products division of International Telephone and Telegraph Co., where he worked on problems associated with frequency translators. Prior to obtaining the Ph.D. degree, Dr. Keblawi was an instructor at North Carolina State University. He is a member of the Institute of Electrical and Electronics Engineers, Phi Kappa Phi, Tau Beta Pi, and Eta Kappa Nu.

HENRY KRESSEL received the B.A. degree from Yeshiva College in 1955. He received the S.M. degree in Applied Physics from Harvard University in 1956, and the M.B.A. and Ph.D. degree from the University of Pennsylvania in 1959 and 1965, respectively. Dr. Kressel joined RCA Electronic Components and Devices, Somerville, N. J., in 1959. He has been instrumental in the research and development of high-frequency mesa, planar, and planar epitaxial silicon transistors. In 1961 Dr. Kressel became head of the microwave diode group, where he was responsible for research, development, and pilot-line engineering of gallium arsenide and silicon varactor diodes. From 1963 to 1965, he was on a leave of absence doing doctoral work in metallurgy. Upon his return to RCA in 1965, he became head of the Device Physics Group in the Technical Programs Laboratory (Somerville, N. J.). He joined RCA Laboratories, Princeton, in 1966. His current research interests are in the area of electro-optical devices and defects in the III-V compounds. Dr. Kessel is a member of the Institute of Electrical and Electronics Engineers and the American Physical Society.

HELENE E. KULSRUD received the B.A. degree magna cum laude from Smith College in 1953 and the M.S. degree in Astrophysics from the University of Chicago in 1955. She was a member of the IBM Data Processing Center from 1954 to 1955. She joined RCA Laboratories in 1957 and has worked on scientific problem solving, computer solution of ordinary and partial differential equations, machine simulation, and digital solution of electron optical problems. She is currently concerned with programming languages for interactive computer use. Mrs. Kulsrud is a member of Phi Beta Kappa and the Association for Computing Machinery. At the present time, she is a member of the Department of Engineering and Applied Science at Yale University.

JOHN A. OLMSTEAD received the B.S. degree in electrical engineering from the University of Buffalo in 1952 and the M.S. degree in electrical engineering in 1957 from Newark College of Engineering. He has also taken graduate courses in mathematics and physics at the University of Buffalo. In 1952 he joined the Radio Corporation of America as a specialized trainee. From 1952 to 1954 he worked on gaseous discharge devices at RCA Laboratories, Princeton. In 1954 he joined Electronic Components and Devices, Harrison, N. J., where he worked on the design and development of gas-filled and other specialized industrial tubes. Between 1950 and 1960 he alternated between assignments at RCA and the University of Buffalo where he was an Assistant Professor of Electrical Engineering. In 1960 he returned to RCA on a full time basis, joining the Semiconductor and Materials Division. He has worked on GaAs solar cells, p-n junction field-effect devices, high-frequency power bipolar transistors and MOS field-effect transistors. He is responsible for the multi-emitter concept basic to the "overlay" transistor technique. Mr. Olmstead is a member of the Institute of Electrical and Electronics Engineers and Sigma Xi.

L. S. ONYSHKEVYCH received a B.E.E. degree from City College of New York in 1955. He worked at the Research Laboratory of Electronics of the Massachusetts Institute of Technology in the area of multiaperture magnetic logic from 1955 to 1957, receiving an M.S. degree in Electrical Engineering in 1957. In 1957 he joined RCA Laboratories where he worked in the areas of magnetic logic and memory systems, parametric oscillators, and tunnel diodes. In 1959 he returned to MIT, and in 1962 he received the E.E. degree. During the period 1961-1963 he worked at the MIT Lincoln Laboratories in the field of thin magnetic film memories. In September 1963 he rejoined RCA Laboratories, where he has been working in the area of sonic film memories. Mr. Onyshkevych is a Member of the Institute of Electrical and Electronic Engineers, Tau Beta Pi, Sigma Xi, and Eta Kappa Nu.

BRUCE B. ROBINSON received his B.S. degree in physics from Drexel Institute of Technology in 1956 and M.A. and Ph.D. degrees from Princeton University in 1958 and 1961, respectively. His doctoral dissertation was in the field of plasma transport theory. He was a research assistant at the University of Chicago and at the University of California where he worked in the areas of plasma transport, plasma stability, and general relativity. Dr. Robinson joined the staff of RCA Laboratories in 1963 and has since that time worked principally in the areas of atomic physics and plasma stability. He is a member of the American Physical Society.

JOSEPH H. SCOTT received the A.B. degree in chemistry from Lincoln University, Pennsylvania. In 1958 he attended graduate school at Howard University, Washington, D.C., and is presently pursuing a degree in Electrical Engineering at Newark College of Engineering. In 1959, he joined RCA Electronic Components and Devices, Somerville, N.J., where he engaged in research and development of semiconductor devices. He has worked on GaAs solar cells, p-n junction unipolar devices, high-frequency high-power bipolar transistors and, most recently, MOS transistors. Together with J. A. Olmstead he did the initial work on multi-emitter "overlay" transistors. In 1966, he joined RCA Laboratories, Princeton, N.J. Mr. Scott is a member of the Electrochemical Society.

R. SHAHBENDER received the B.E.E. degree from Cairo University, Cairo, Egypt, in 1946, the M.S. degree in Electrical Engineering from Washington University, St. Louis, Mo., in 1949, and the Ph.D. degree in Electrical Engineering from the University of Illinois in 1951. After completing his undergraduate studies, Dr. Shahbender joined the Anglo-Egyptian Oilfields, Ltd. for a period of two years and did work in geophysical oil exploration. In 1951 he joined the staff of the Minneapolis-Honeywell Regulator Co., Brown Instruments Division, and conducted research in the behavior of nonlinear control systems. In 1955, Dr. Shahbender joined RCA in Camden, N. J., and worked in the areas of adaptive systems, nonlinear filters, electron-beam devices, ultrasonic devices and airborne fire control systems. Dr. Shahbender transferred to RCA Laboratories, Princeton, N. J., in 1959 and has been active in the area of high-speed digital memory systems. He is presently head of the magnetic memory and devices research group in the Computer Research Laboratory. Dr. Shahbender is a member of the Institute of Radio Engineers, the American Institute of Electrical Engineers, the Franklin Institute, and the Societies of Sigma Xi, and Eta Kappa Nu.

HILLEL WEINSTEIN received a B.E.E. degree from New York University, and an M.S. degree and, in 1965, a Ph.D. from the Polytechnic Institute of Brooklyn. He was a member of the Technical Staff at RCA Laboratories from 1960 to 1966. He is presently at Xerox Corporation, Webster, New York. Dr. Weinstein has been active in the areas of high-speed switching circuits, optoelectronic image processing, and logic design. In the area of high-speed switching circuits, he conceived and analyzed a time-variant system for pulse stretching. In the area of image processing, he worked on the problem of spatially shifting an optical image without recourse to mechanical motion. His most recent work involved logic conception and design of multiple-match detection and sequential read-out systems for a magnetic content addressable memory. Dr. Weinstein is a Member of Eta Kappa Nu, Sigma Xi, and Tau Beta Pi.

KARL H. ZAININGER received the B.E.E. degree (magna cum laude) from City College of New York in 1959; the M.S.E. in 1961, the M.A. in 1962, and the Ph.D. in Engineering Physics from Princeton University. In 1959 Dr Zaininger joined the staff of RCA Laboratories where he has been engaged in research on various semiconductor devices. He has been involved in research on Silicon based MOS devices since their original inception. His research activities with the MOS devices have included basic studies in the area of semiconductor surface physics. He is presently concerned with MIS device physics and technology, with measurement techniques, and with the physics of radiation damage in MIS systems. Dr. Zaininger is a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, the Institute of Electrical and Electronics Engineers, and the Shevchenko Scientific Society.