RCA
Review

RCA REVIEW

a technical journal

VOLUME 29   SEPTEMBER 1968   NUMBER 3

CONTENTS

RCA REVIEW

MOS FIELD-EFFECT TRANSISTOR DRIVERS
FOR LAMINATED-FERRITE MEMORIES

BY

W. A. BÖSENBERG, D. FLATLEY, AND J. T. WALLMARK*

RCA Laboratories
Princeton, N. J.

Summary—Various alternatives for driving a low-current ferrite memory with a cycle time of 2 μsec are considered. One approach, based on MOS field-effect-transistor drivers, is selected for experimental tests because of complete bidirectionality, high breakdown voltage, and simple technology. The drivers were fabricated in 64-unit monolithic strips on 10-mil centers with a channel length of 0.3 mil. The channel width was 110 mils for the word drivers and 30 mils for the digit drivers.

The word drivers delivered 120 mA with a source-to-drain voltage of 4 volts, a substrate bias of −9 volts and a gate swing of 4 volts. The digit drivers, essentially smaller copies of the word drivers, delivered 20 mA under similar circumstances. The gate capacitance was 31 pF per word driver and the drain capacitance with −9 volts substrate bias was 14 pF per word driver.

On the basis of limited post-fabrication testing and limited life testing, it was found that a practical size for the word driver strips, compatible with an experimental yield of 30%, was 32 drivers to a monolithic strip. The limitation in size was primarily caused by the relatively large gate area, demanding silicon dioxide with a very low density of dielectric defects.

CONSIDERATIONS OF MOS TRANSISTORS FOR CURRENT DRIVERS

THE APPLICATION to ferrite memories of methods of simultaneous fabrication of large numbers of elements and their interconnections has led to monolithic laminated ferrite memories.[1,2] In these memories, the magnetic cores are replaced by ferrite sheets, so that, in a manner of speaking, the cores are all processed together and need not be handled individually. For this reason smaller equivalent dimensions, and therefore smaller currents, are possible than in the case of conventional ferrite cores. Thus, with the ferrite sheet memory, it is feasible to use MOS field-effect transistors for switching the currents need to drive the memory. With the cores, bipolar transistors with their much lower impedance levels provide a

[1] I. Gordon, R. L. Harvey, H. I. Moss, A. D. Robbi, J. W. Tuska, J. T. Wallmark and C. Wentworth, "An MOS-Transistor-Driven Laminated Ferrite Memory," *RCA Review*, Vol. 29, p. 199, June 1968.

[2] R. Shahbender, "Laminated Ferrite Memories—Review and Evaluation," *RCA Review*, Vol. 29, p. 180, June 1968.

* J. T. Wallmark is presently at the Chalmers Institute of Technology, Göteborg, Sweden.

better match and are universally used. The various considerations that enter into a choice of driving circuits may be summarized as follows:

The bidirectionality of currents is a problem in regular diffused-base n-p-n silicon transistors, since the collector region has a smaller carrier concentration than the base region and, therefore, the inverse common-emitter current gain is low. A pair of n-p-n transistors can be used, with the emitter of one transistor connected to the collector of the other, and vice-versa, and with some base resistance added to avoid "current hogging." This arrangement works well, but the maximum voltage is limited to the emitter breakdown voltage. This is usually around 7 volts but it can be increased to about 15 volts by reducing the base impurity concentration (and the frequency response) somewhat. In the present application, a minimum breakdown voltage of 25 volts is required because laminated ferrite arrays have a comparatively large back voltage. This large back voltage results because the magnetic material is distributed evenly between, as well as in, the core positions. Thus, pairs of diffused-base n-p-n transistors cannot be used.

Alloyed-emitter/alloyed-collector structures have good inverse current gain. They require individual etching, however, and do not lend themselves to batch processing. For this reason, alloyed silicon transistors have found application only in the high-power, low-frequency single-device area.

Integrated circuit techniques use diffused-base, diffused-emitter, n-p-n transistors almost exclusively. p-n-p transistors are usually avoided, partly because of the lower performance (lower current gain for equal dimensions) caused by the low mobility of holes compared to electrons, and partly because it is difficult to fabricate n-type base regions free of channel formation (which leads to poor reverse characteristics of back-biased p-type collectors). Use of a p-n-p transistor with a low current gain followed by an n-p-n transistor with high current gain provides sufficient current gain, but the frequency response is inadequate. In addition, the fabrication of both p-n-p and n-p-n transistors on the same substrate, yet isolated, is cumbersome. Therefore, a complementary bipolar driver approach was ruled out for this project.

MOS field-effect transistors, on the other hand, can easily be fabricated with a breakdown voltage $\geq 25V$, with pulse rise and fall time ≤ 200 nsec, and with complete bidirectionality. The technology involved is relatively simple. With field-effect transistors, only one transistor per driver (and no passive components) is needed. A 64-output

integrated strip, therefore, contains only 64 field-effect transistors; all sources are common but the gate and drain of each unit is individually accessible.

In addition to the word-current drivers, the laminated ferrite memory system, shown in Figure 1, contained digit-current drivers and
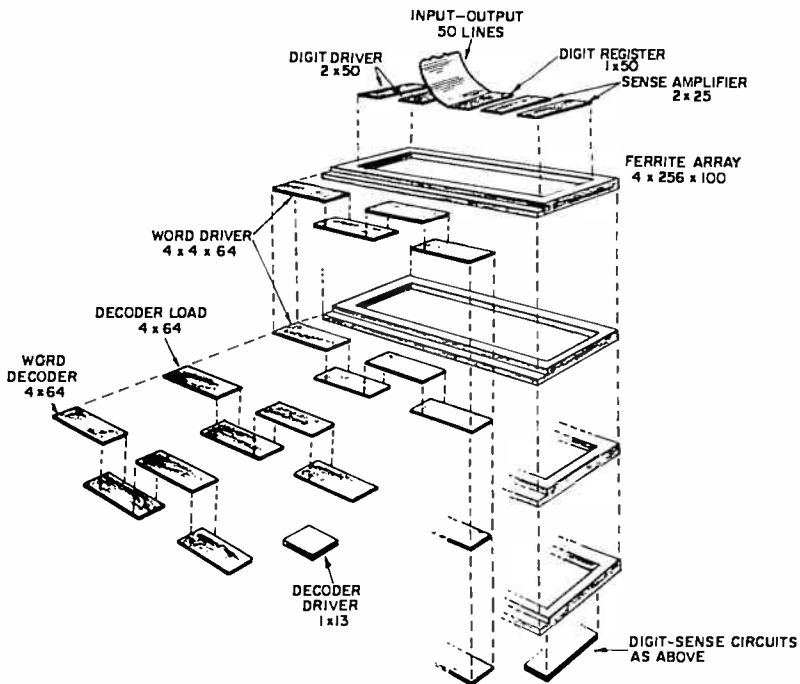
Fig. 1—Exploded view of the laminated ferrite memory sub-system. Four ferrite planes, each containing 256 × 100 lines, are each driven by four word-driver strips of 64 MOS transistors each. The decoder and digit-driver strips shown are shared by the four ferrite frames.

word-decoders, all fabricated by the same technique, i.e., all employed MOS transistors exclusively.[3] The word decoder has been described elsewhere.[4] The digit drivers are essentially small copies of the word drivers, reflecting the fact that the digit drive currents are 20 mA as compared to word-drive currents of 120 mA.

---

[3] J. T. Wallmark, W. A. Bösenberg, E. C. Ross, D. Flatley, and H. Parker, "MOS Field-Effect Transistor Technology," Final Report, Contract NAS 1-5794, National Aeronautics and Space Administration, Wash., D. C., August 1967.

[4] J. T. Grabowski, "A 64-Output MOS Selection Tree Using Large-Scale-Integration Technique," *RCA Review*, Vol. 29, p. 324, Sept. 1968.

DESIGN CONSIDERATIONS

MOS transistors with full gate strip structures (see Figure 4) have bidirectional current properties and the source and drain contacts can be interchanged. An MOS transistor can be designed either with a built-in (depletion type) or an induced channel (enhancement type), i.e., for zero gate bias, there is or is not a drain current flowing. For this application, an enhancement unit with no drain current at zero gate bias would be desirable. However, as a result of the processing, particularly the double-gate insulator construction needed for long-
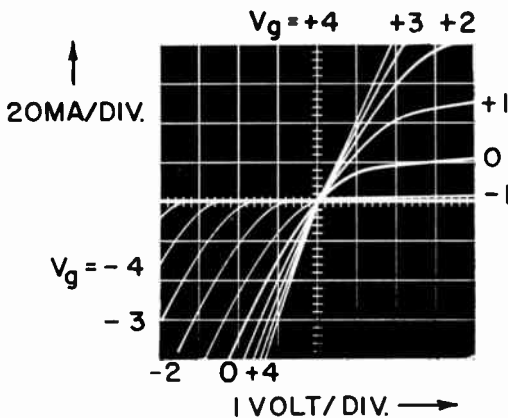
Fig. 2—Drain current versus drain voltage for an MOS driver transistor. Gate voltage is parameter.

term stability under bias, the transistors were depletion type, requiring about 3 volts negative bias to reduce the current to less than 100 μA per unit on all units.

Note that quite symmetrical drain-current/drain-voltage characteristics occur if the same potential (either positive or negative) is applied to both the gate and drain (Figure 2). In this mode of operation, the sources are grounded. The systems requirements are such that larger drain currents are needed for reading than for writing. Therefore, the MOS transistors are used with negative bias in the read mode and positive bias in the write mode.

The drain current I_D in the saturation region can be calculated from first-order theory;[5]

---

[5] J. T. Wallmark and H. Johnson, eds., *Field Effect Transistors, Physics, Technology and Applications*, Prentice-Hall, Inc., Englewood Cliffs, 1966.

$$I_p = \frac{\mu_{\text{eff}}\epsilon_{ox}}{2t d_{ox}} W (V_g - V_p)^2$$

where

- μ_{eff} is the effective majority carrier mobility in the channel region,
- ϵ_{ox} is the dielectric constant of the gate insulator,
- d_{ox} is the thickness of the gate insulator,
- W is the channel width,
- t is the channel length,
- V_g is the gate voltage,
- V_p is the pinch-off voltage.

It can be seen that for large drain currents (100 to 150 mA are required), the aspect ratio of width to length (in the direction of current flow) should be maximized while the thickness of the insulator should be minimized. The latter can be done only to a point where the field across the gate insulator equals the maximum dielectric field (about 10^7 V/cm for this silicon dioxide layer). In practice, gate breakdown voltages of larger than 20 volts require at least 500 Å of silicon insulator. A channel length of 0.3 mil and a channel width of slightly more than 100 mils were chosen for the word drivers. There is a high field at the drain contact,[3] which limits the source–drain voltage to about 25 to 35 volts in these devices. For the digit devices, the channel width of 28 mils corresponding to a maximum current of 20 mA was chosen while the channel length and oxide thickness were the same as for the word drivers.

Since the laminated ferrite has orthogonal lines on approximately 10-mil centers, the MOS drivers were designed on the same center-to-center spacing. Both the drain and the source are U-shaped with the gate electrode meandering between them (see Figures 3, 4, and 5). All source contacts are combined as a common bus. The drain contacts go to one end of the silicon chip, while the gate electrodes pass underneath the dielectrically isolated source bus to the other side of the silicon chip.

For good yield, high utilization of the silicon surface area is necessary. The geometry in Figure 6a, which was used for the driver strips, resulted from a compromise in mask making. Originally, a geometry as shown in Figure 6b was used. At that time satisfactory masks with consistent 0.3 mil channel lengths could not be obtained, even with the use of a recticle layout for the step and repeat process that split the

artwork along the center of the source area instead of the center of the much smaller gate area. Therefore, the perimeter of the driver strips was extended by 33%, since only three edges of the two drain contacts draw current (Figure 6a) instead of all edges of drain contacts drawing current (Figure 6b). At present, advances in photomask technology would make the original layout feasible and would reduce both the gate and the drain capacitances by 25%. Also the yield figures should improve considerably.
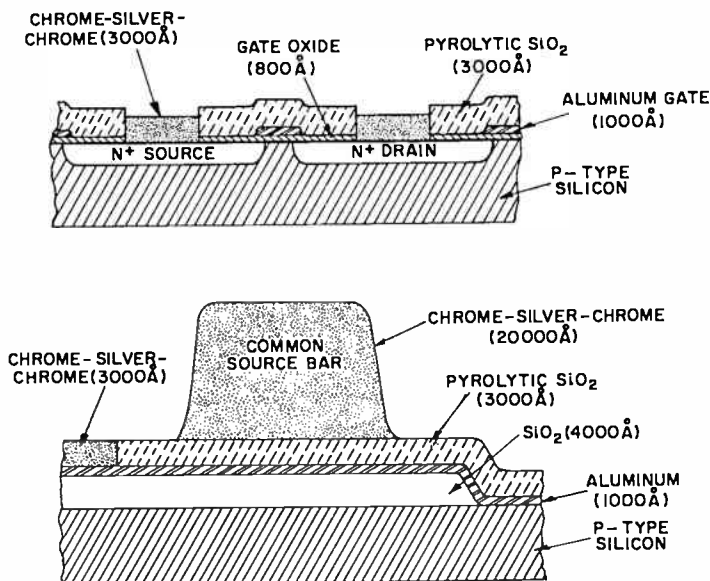
Fig. 3—Cross-section of MOS driver transistor.

The gate capacitance for a strip width of 0.5 mil and 800 Å insulation was measured as 31 pF with the gate biased into the accumulation region, which is the maximum value. The drain capacitance at zero substrate bias is quite large, about 160 pF. For this reason, the substrate is biased negatively in operation. At a substrate voltage of −9 volts and a drain voltage of +6 volts, the drain capacitance is 14 pF.

The precision photolithographic masks contain 64 word drivers with 110-mil channel width and, on a different set, 100 digit drivers with 30-mil channel width.

A comparison of the complexity of the driver circuits for the laminated-ferrite memory system is shown in Table I. The table gives figures on the total gate area per strip, which is related to the proba-
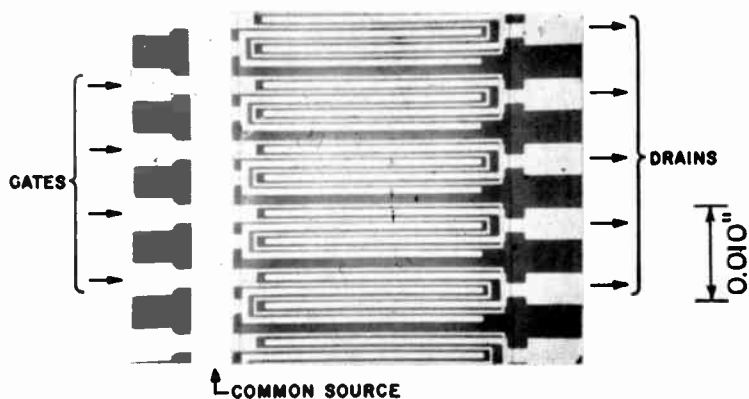
Fig. 4—Top view of part of the word-driver strip.

bility of gate shorts through defects in the gate insulation; the total device area per strip, which is related to yield; the number of devices per monolith; and the number of crossovers per monolith, which is related to the probability of short circuits in the lines. As can be seen, the word driver strip and, particularly, the decoder are more complex than judged commercially feasible at the present time.
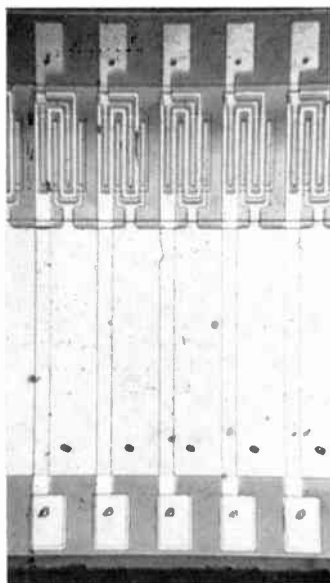
Fig. 5—Top view of part of the digit driver strip.

Table I—Comparison of Silicon Switches.

	Total Gate Area Per Strip (mil²)	Total Device Area per Strip (mil²)	Number of Devices Per Strip	Number of Crossovers Per Strip
Word switch strip	520	45000	64	64
Digit switch strip	230	84000	100	100
Decoder strip	2200	130000	85	1634
Decoder load strip	150	54000	64	64

PROCESSING

The substrate was made of 10 ohm-cm n-type Czochralski-grown silicon single crystal of (111) or (100) orientation, chemically polished. Source and drain contacts were phosphorus diffused, with a surface concentration of 5×10^{20} cm^{-3}. The gate oxide was grown in two steps. A dry oxide layer of about 600 Å was grown first; this was then covered with a 200-Å-thick phosphor-silicate layer.

A number of metallization schemes were tried. Aluminum contacts are most straightforward, but do not allow soldering or welding of the driver strips to the laminated ferrite. Therefore, aluminum was used only on the gate contacts; chrome–silver–chrome metallization was adopted for the source/drain contacts, with blending during the evaporation.[6] Dielectric isolation using 1000 Å of pyrolytically de-
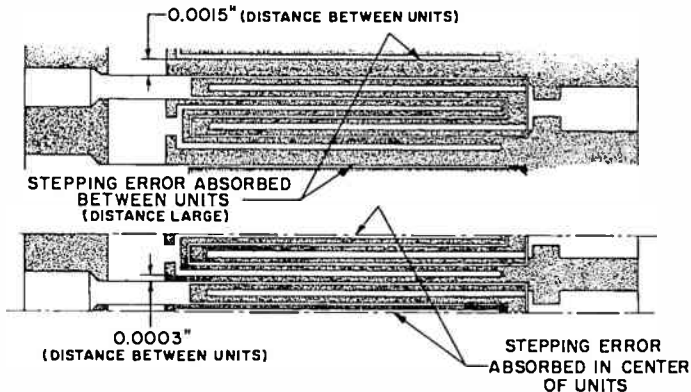
Fig. 6—Step-and-repeat pattern for MOS driver transistor: (top) stepping one transistor pattern (errors are taken up between units) and (bottom) stepping two halves of a transistor pattern (errors are taken up in source bar).

[6] E. C. Ross and J. T. Wallmark, "Extending the Life of Chromium-Silver Metallization on Silicon Devices," Proc. 1967 Annual Symposium on Reliability Physics, Los Angeles, 1968.

posited silicon dioxide was used for the common-source bus bar over the gates. The bus represents a series resistance to the drain current and should have as high conductivity as possible ($\approx 20,000$ Å thickness). Therefore, the bus was re-inforced by an additional evaporation through a metal mask. In addition, contacts to the bus were provided at both ends, further reducing the series resistance that amounted to about 1 ohm (corresponding to a voltage drop of 0.1 to 0.2 volt).

On the digit device strip, all digits would be activated simultaneously. Therefore, the source bus was made 38 mils wide, resulting in a series resistance of about 0.1 ohm and a voltage drop of about 0.2 V.
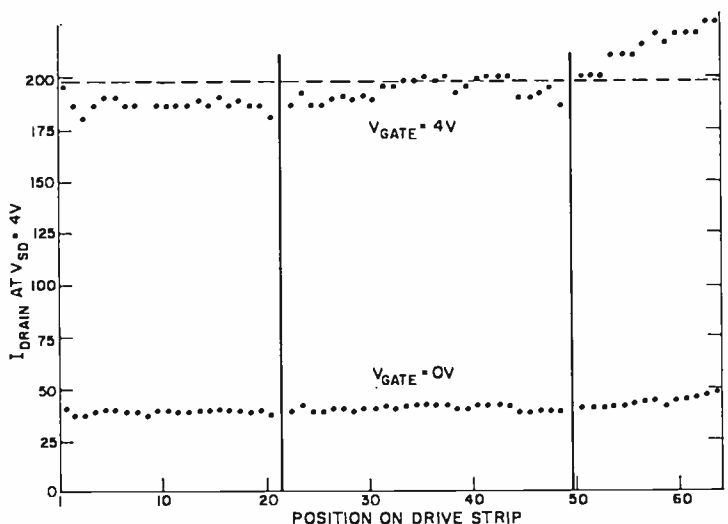
Fig 7—Drain current at $V_{sn} = 4$ V, $V_n = 0$ V and 4 V for typical driver strip. Two units indicated by vertical bars have gate shorts.

DEVICE YIELD IN FABRICATION AND ELECTRICAL PERFORMANCE

A 64-unit word-driver strip occupies a silicon area of approximately 70×650 mils. This is a relatively large area for defect-free devices. However, half of this area is occupied by relatively noncritical contact areas. With limited laboratory fabrication facilities (without clean laminar flow hoods), only one strip with all 64 units good has been obtained. However, the number of half strips with 32 consecutive good units on two typical wafers, shown in Figure 7, was 13 (out of a total of 40) about 30% yield. However, the strips were not tested at the voltage extremes encountered in an operating system. With clean-room facilities, higher yield may be expected, but it is still felt

that a 32-unit driver strip is a practical compromise as regards size versus yield.

The drain currents on a given wafer were quite uniform. Measurement results with a typical word-driver strip are shown in Figure 7. At a gate voltage of $V_g = 4$ V and a source-drain voltage of $V_{sd} = 4$ V with the substrate connected to the source, the average drain current is close to 200 mA. For our application, only 100-150 mA drain current is required, which can be achieved with a gate bias of 2-3 volts. The source–drain current is approximately 40 mA at zero gate bias and can be reduced to less than 100 μA with negative gate bias.

A map of the completed wafer surface showing the zero-gate-bias drain current of two full wafers is shown in Figure 8. The number of good units for each 64-unit strip is indicated, together with the maximum spread in drain current for the strip. There was a systematic variation of the drain current for each wafer, as indicated by the equi-current contours. The reason for the current variation is related to the difference in processing experienced by various parts of the wafer. In addition Figure 8(a) shows an area of high series resistance in source–drain contacts resulting in anomalously low drain current. This resistance was probably caused by an oxide layer at the source and drain contacts.

While it was realized that a practical system would require hermetic encapsulation by a surface cover, such as silicon nitride, no attempt was made to develop such a cover. All the tests were done with bare units or, where longer life was essential, with units divided into small groups and encapsulated in TO-5 enclosures. Consequently, life tests and changes in characteristics with life were not studied. It may be of interest to note that the plastic used for the mounting frames contained an ingredient (most likely the hardener, which was an aliphatic amine) that over a period of several months severely degraded the breakdown strength of the silicon dioxide. This degradation was particularly noticeable in units stored in closed containers where the partial pressure of the hardener could build up. For this reason, Photoform glass frames were substituted for the plastic, with the connection fingers brazed to metallized strips.

From destructive tests, it appears that the short-circuited gates, which constituted one of the most prevalent faults, were not caused by faults in the cross-overs but, as in 20 out of 20 cases investigated, by faults in the gate thermal oxide.

The experimental testing of the drivers in combination with a laminated ferrite memory plane with a cycle time of 2 μsec has been described elsewhere.[1]
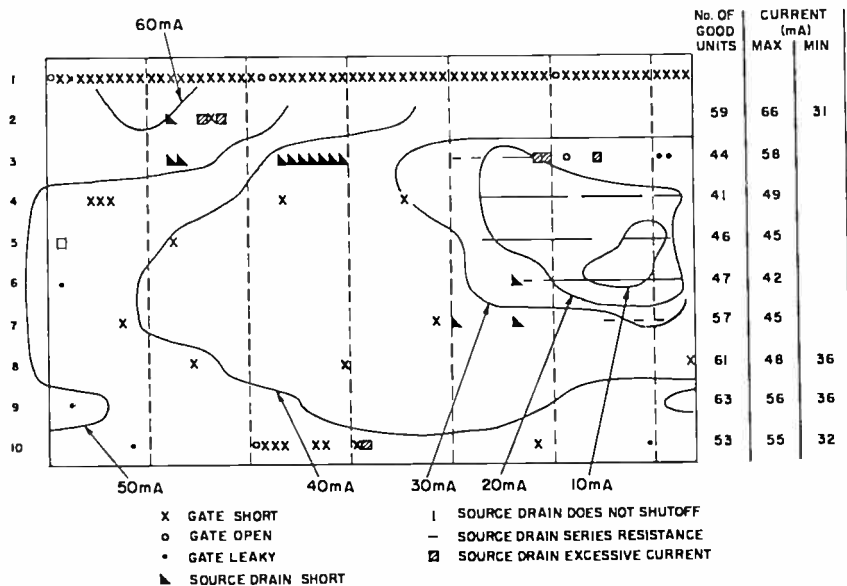
(a)
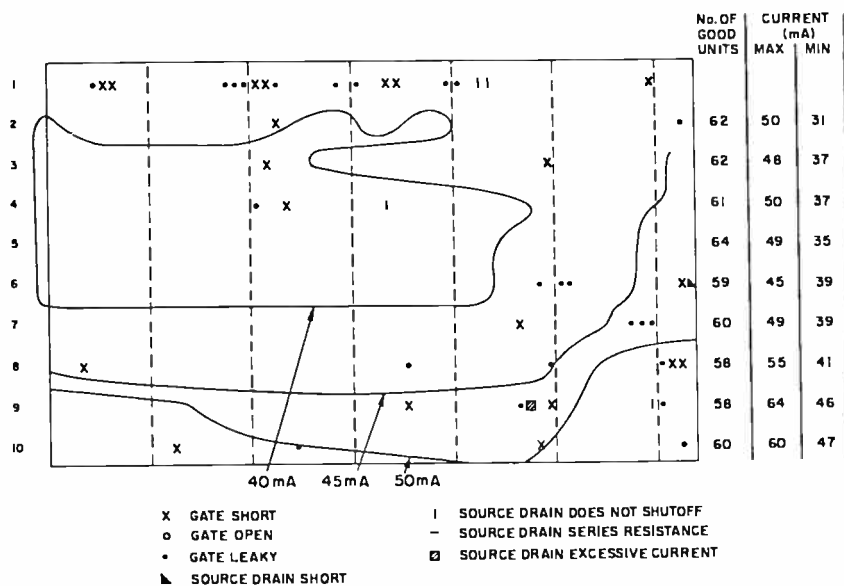
(b)

Fig. 8—Experimental results obtained with two wafers, each containing ten 64-unit driver strips. Contours show equal drain current at $V_{sD} = 4$ V, $V_G = 0$ V. Number of good units per strip and the maximum and minimum drain current for each strip are given to the right.

CONNECTION OF DRIVER STRIPS

The method for interconnection developed for the laminated-ferrite memory used a solder-reflow technique for connecting the various parts of the driving circuitry to themselves and to the ferrite matrix. Each word-driver strip was mounted in a plastic frame as shown in Figure 9.
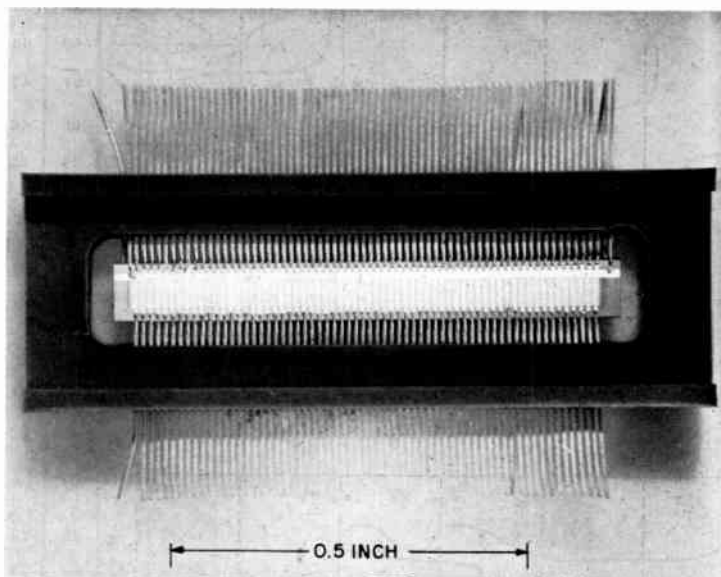
Fig. 9—Word driver strip mounted in plastic frame by solder reflow method. Center-to-center distance 10 mils; frame length 940 mils.

The mounting was done at room temperature except for a heat pulse to about 250° to 300°C of about a millisecond duration on the contact pad, and with very small mechanical pressure. Therefore, the method was expected to introduce no change in the transistor characteristics, and none was observed.

CONCLUSIONS

It has been shown that the bidirectionality of MOS field-effect transistors can be utilized to construct very simple all-MOS drivers for low-current ferrite memories with a cycle time of 2 μsec. On the basis of limited post-fabrication testing and limited life testing, it was found that a practical size of the word driver strips with an experimental yield of 30%, was 32 drivers to a monolithic strip. The limita-

tion in size was primarily caused by the relatively large gate area, which required silicon dioxide with a low-density of dielectric defects.

Because of this limitation, another alternative for the driving circuits, which has appeared more recenly, appears advantageous. This alternative utilizes n-p-n transistors in combination with a small MOS field-effect transistor to reverse the polarity of current (BIMOS).[7] It also has the advantage of simple fabrication in that the MOS transistor, which is a p-channel unit, is obtained with the same processing as that used for the n-p-n transistors. The gate area is very small, even though the total silicon area is comparable to the all MOS approach.

ACKNOWLEDGMENT

---

[7] S. Katz, A. D. Robbi, C. Wentworth, "10^8-Bit Low Power Laminated-Ferrite Memory," National Aeronautics and Space Administration, Contract NAS 12-574.

DEVELOPMENT OF A 64-OUTPUT MOS
TRANSISTOR SELECTION TREE

By

J. T. GRABOWSKI

RCA Laboratories
Princeton, N. J.

Summary—A 64-output selection tree consisting exclusively of p-channel MOS enhancement transistors is described. Possible areas of application include display and information storage systems. The design parameters, fabrication techniques and electrical testing of the tree network are discussed. A complementary decoding circuit incorporating the selection tree operates at a 2 MHz decoding rate with a power dissipation of 20 mW.

INTRODUCTION

A SELECTION tree provides a basic decoding function such that a signal appears on one of N output lines during the period of an applied input code. A 1024-output decoding circuit was required for the word-selection portion of the laminated-ferrite memory.[1,2] The 1024-output circuit was partitioned into 64-output units. This approach reduces the semiconductor wafer to manageable fabrication size, and improves overall yield by permitting selection of only good 64-output selection trees for use in the decoder. The design, fabrication, and evaluation of the 64-output selection tree using MOS transistors are described in the present paper. The MOS transistor is a field-effect, unipolar device with an insulated gate. This device has been extensively covered in recent literature.[3-5] The principal advantages of the MOS transistor, as compared to the conventional bipolar transistor, are extremely high input resistance, ease of fabrication, and improved thermal stability.

---

[1] I. Gordon, R. L. Harvey, H. I. Moss, A. D. Robbi, J. W. Tuska, J. T. Wallmark, and C. Wentworth, "An MOS-Transistor-Driven Laminated-Ferrite Memory," *RCA Review*, Vol. 29, p. 199, June 1968.

[2] J. Grabowski, *An MOS Word Address Decoder*, Master's Thesis, University of Pennsylvania, May 1966.

[3] H. K. J. Ihantola and J. L. Moll, "Design Theory of a Surface Field-Effect Transistor," *Solid State Electronics*, Vol. 7, p. 423, 1964.

[4] C. T. Sah, "Characteristics of the Metal-Oxide-Semiconductor Transistors," *IEEE Trans. El. Dev.*, p. 324, July 1964.

[5] J. T. Wallmark and H. Johnson, *Field-Effect Transistors*, Prentice-Hall, Englewood Cliffs, N. J., 1966, pp. 135-144.

<div style="text-align:center">SELECTION-TREE DESIGN</div>

Logic Configuration

The selection tree must provide output signals on the order of $+10$ volts to drive the word-selection switches.[1] Since each MOS transistor
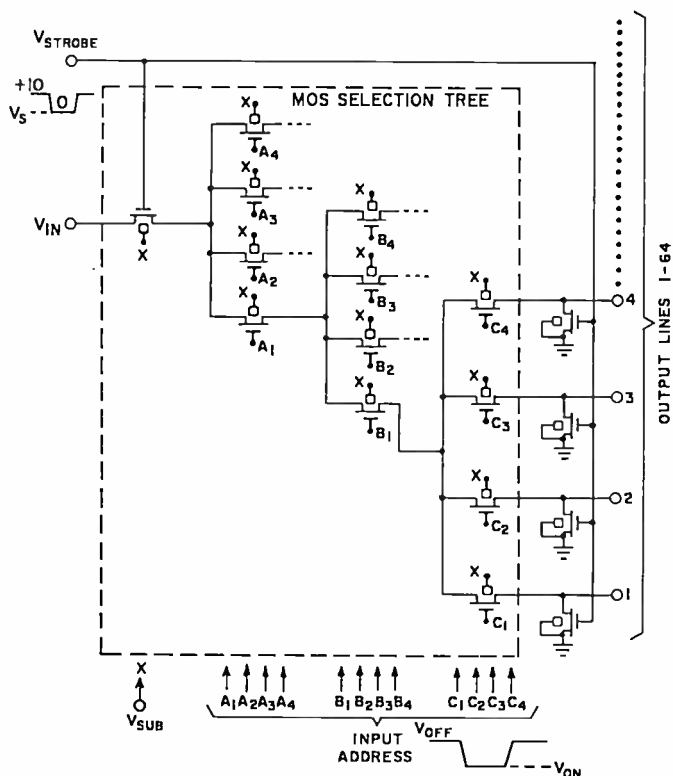
Fig. 1—Complementary decoder incorporating MOS selection tree.

can be turned ON or OFF by controlling its gate voltage, the most economical logic array for a decoder is the tree-type structure illustrated in Figure 1. An ON output appears only at the particular output line selected by the input code; the other output lines remain in the OFF state.

The internal coding system of the decoder (binary, quaternary, etc.) determines the number of individual transistor units needed, the number of levels of the decoder, and the intraconnection complexity. Each transistor stage in a signal chain leading to an output line repre-

sents one level. The important features of the basic tree-type selection
network using binary, quaternary, and octal codes are compared in
Table I for a 64-output decoder.

Table I—Characteristics of Selection Trees Using Different Internal
Coding Systems for a 64-Output Decoder

Internal Code	Requirements			
	MOS Transistors	Levels	Input Code Lines	Signal Crossover Points
Binary	126	6	12	252
Quaternary	84	3	12	336
Octal	72	2	16	576

In general, the number of MOS transistors and the number of
levels decrease for higher-order internal coding systems, but the num-
ber of input code lines and the intraconnection problem due to cross-
overs increase. The selection of an internal coding system for the de-
coder must, therefore, be a compromise, with any increase in decoder
performance weighed against the intraconnection problem and the
problem of converting from the external binary code (i.e., the code
used in the computer system external to the decoder) to the selected
internal code. The present work considers only the quaternary code.

Selection Tree Model

A mathematical model, consisting of a discrete G-C (conductance-
capacitance) transmission line with voltage-dependent conductances
and capacitances was used to characterize the MOS selection tree op-
erating as a decoder. The validity of the model was established by
comparison of predicted and actual operating performance of a 64-
output decoding circuit using discrete MOS transistors.[2] The polarity
and physical dimensions of the MOS transistors can be varied within
the model to study their effect on such performance factors as output
signal rise time and uniformity.

The specific model for a 64-output selection tree, with the output
lines terminated in load conductances(G_4) to operate as a decoder, is
shown in Figure 2. Only the conductances in the ON signal chain are
considered since all OFF transistors are assumed to have zero con-
ductance. The equations governing the transient behavior of the model
in Figure 2 are

$$C_1 \frac{dV_1}{dt} = I_{IN} - G_1(V_1 - V_2) ,$$ (1a)

$$C_2 \frac{dV_2}{dt} = G_1(V_1 - V_2) - G_2(V_2 - V_3) \ , \tag{1b}$$

$$C_3 \frac{dV_3}{dt} = G_2(V_2 - V_3) - G_3(V_3 - V_4) \ , \tag{1c}$$

$$C_4 \frac{dV_4}{dt} = G_3(V_3 - V_4) - G_4 V_4 \ . \tag{1d}$$
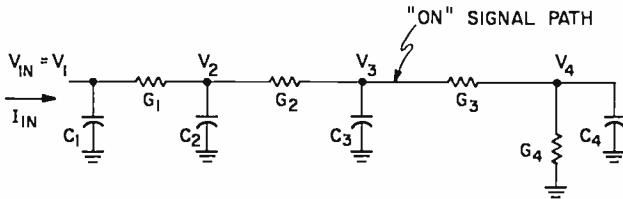
Fig. 2—Model to calculate performance of 64-output MOS decoder.

Conductance Characteristics

The MOS transistor may be operated in either the saturated or non-saturated mode. In the non-saturated, or linear, region, the drain-to-source current (I) is an increasing function of the drain-to-source voltage (V_{DS}). In the saturated region, this current is independent of changes in the drain-to-source voltage, neglecting the second-order effects.[5]

The equations that govern the channel conductance of an MOS transistor over a wide range of gate fields, V_{GS}/T_{ox}, have been established analytically[6] and verified experimentally by testing n-type MOS transistors of varying sizes. All equations in the present section refer to n-type MOS transistors. The equations characterizing p-type MOS transistors are identical in form to those for n-type, but the polarities of the gate and substrate voltages are reversed. For p-type transistors, V_{SG} and V_{SUBS} replace V_{GS} and V_{SSUB} in the device equations as positive quantities. I then refers to source-to-drain current.

---

[6] See Ref. (5), pp. 144-148.

ON transistor, linear region: $(V_{GS} - V_T) \geqq V_{DS} > 0$

$$G = \frac{\mu_o \epsilon_{ox}}{T_{ox}} \frac{W}{L} V_{GC} \left[1 - \frac{V_{GC}}{V_{DS}\left(1 + \dfrac{V_{GS} - V_T - V_{DS}}{V_{GC}}\right)} \ln \frac{\left(1 + \dfrac{V_{GS} - V_T}{V_{GC}}\right)}{} \right], \quad (2a)$$

ON transistor, saturated region: $V_{DS} > (V_{GS} - V_T) > 0$

$$G = \frac{\mu_o \epsilon_{ox}}{T_{ox}} \frac{W}{L} V_{GC} \left[\frac{V_{GS} - V_T}{V_{DS}} - \frac{V_{GC}}{V_{DS}} \ln \left(1 + \frac{V_{GS} - V_T}{V_{GC}} \right) \right], \quad (2b)$$

OFF transistor: $(V_{GS} - V_T) \leqq 0$

$$G = 0 , \qquad\qquad (2c)$$

where G = total channel conductance (I/V_{DS})

μ_o = mobility of the majority carriers in the conduction channel neglecting the effects of diffuse surface scattering,

ϵ_{ox} = dielectric constant ($3.8\ \epsilon_o$) of the oxide insulator between the metal gate and the semiconductor,

ϵ_{Si} = dielectric constant ($11.8\ \epsilon_o$) of the semiconductor,

T_{ox} = oxide insulator thickness between metal and semiconductor,

W = channel width in the direction perpendicular to drain–source current flow

L = channel length in direction of current flow (usually $W > L$),

V_{GS} = gate-to-source voltage,

V_T = gate-to-source threshold, or pinch-off, voltage,

and V_{GC} is defined as

$$V_{GC} = \frac{\epsilon_{Si}}{\epsilon_{ox}} T_{ox} E_{SC} , \qquad\qquad (2(d))$$

where E_{SC} = gate field at which the diffuse surface scattering mobility becomes equal to μ_o.

Equations (2) account for the diffuse surface scattering of majority carriers in the conduction channel, which becomes important at high gate fields. The general form of the conductance characteristic as a function of V_{GS} is illustrated in Figure 3, including the distinction between the linear and saturated regions. Parasitic resistance in the

source and drain contacts can also produce a "saturation" of the channel conductance characteristic similar to that shown in Figure 3. The present work assumes this saturation is predominantly a result of diffuse surface scattering at high gate fields, and uses Equations (2) to characterize the channel conductance. Average values for μ_o and E_{SC} have been derived from measurements on the previously mentioned n-type MOS transistors of varying geometries:

$$\mu_o = 357 \ cm^2/volt\text{-}sec,$$

$$E_{SC} = 4.9 \times 10^5 \ volts/cm.$$
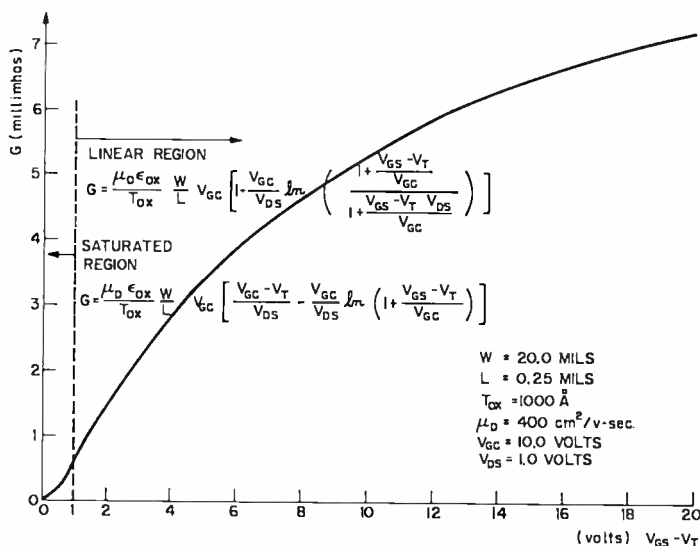
Fig. 3—Conductance characteristic for MOS transistor.

The threshold voltage of an MOS transistor is determined by the electric charge distribution in the vicinity of the conduction channel. This threshold voltage can be changed by altering the substrate doping level, or the source-to-substrate voltage V_{SSUB}. The magnitude of the change in threshold voltage from its $V_{SSUB} = 0$ value (V_{T0}) has been predicted by Hofstein for the case of an abrupt junction between the source region and the substrate:[7]

$$\Delta V_T = 2 \frac{\epsilon_{Si}}{\epsilon_{ox}} \left(\frac{N q T_{ox}^2}{\epsilon_{Si}} \right)^{1/2} \left[(V_{SSUB} + V_0)^{1/2} - V_0^{1/2} \right], \qquad (3)$$

---

[7] See Ref. (5), pp. 150-152.

where V_0 is the semiconductor junction barrier voltage. Experimental measurements of $\Delta V_T = \Delta V_T(V_{SSUB})$ for n-type MOS transistor units indicate that

$$\Delta V_T = 0.70 \ (V_{SSUB} + 0.8)^{0.59} \text{ for } V_{SSUB} \geqq 4.0 \text{ V} \tag{4}$$

provides a good description of the pinch-off voltage variation as a function of source-to-substrate voltage for 10 ohm-cm substrate material.
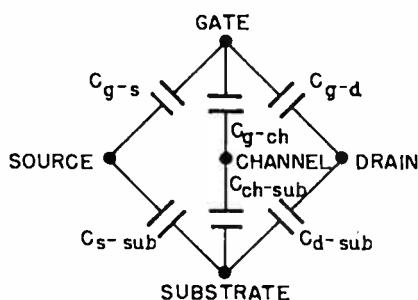
Fig. 4—Characterization of MOS transistor capacitances.

Capacitance Characteristics

The capacitance associated with the MOS transistor can be represented by the model illustrated in Figure 4. The $C_{ch\text{-}sub}$ capacitance, usually an order of magnitude less than the C_{g-ch} capacitance, can be neglected for all practical purposes because it is so small. Studies on typical transistor units resulted in the following expressions for the remaining capacitances:

$$C_{g-ch} = \frac{\epsilon_{ox}LW}{T_{ox}}, \quad \text{for } (V_{GS} - V_T) > V_{DS}, \tag{5(a)}$$

$$C_{g-s} = \frac{\epsilon_{ox}L_oW}{T_{ox}}. \tag{5(b)}$$

$$C_{s-sub} = \frac{0.67 \ A_s}{(V_{SSUB} + 0.8)^{0.4}}, \tag{5(c)}$$

where A_s is the area of the source region in contact with the substrate and L_0 is the length of the metal gate overlap on either a source or drain region. The capacitances C_{g-d} and C_{d-sub} may be obtained from the last two expressions by substituting the drain for the source in the subscript notation. When the transistor is ON in the linear region,

Fig. 5—Partial circuit structure of a 64-output all-n-type MOS decoder using quaternary internal code organization.

$(V_{GS} - V_T) > V_{DS}$, then C_{g-s} and C_{g-d} appear in parallel with C_{g-ch} through a large conductance; in the OFF state, $(V_{GS} - V_T) < 0$, C_{g-ch} is essentially isolated from C_{g-s} and C_{g-d}.

Comparison of Selection Trees

The preceding G-C transmission line model was used to compare the performance of all-n-type and all-p-type 1024-output selection trees operating as decoders. The output lines of the selection trees were terminated in a nonlinear MOS load conductance of the same tree po-

larity as illustrated in Figure 5. An analysis of the 1024-output decoding circuit was chosen to emphasize the performance differences among selection trees and also to indicate the general level of realizable performance from an actual circuit.

In the performance analysis, all the MOS transistors in the selection tree of a particular decoder consist of identical, interdigitated structures. The dimensions of the individual source/drain are $W/2 \times 2.0$ mils. Imposition of this identical geometry condition simplifies both the decoder design and fabrication process. A few decoders are analyzed to determine if their performance can be significantly improved by increasing the channel width of the MOS transistors near the input end of the selection tree.

Values for the following constants are necessary if the electrical properties of the MOS transistors in the decoder are to be completely characterized by Equations (2), (4), and (5): (1) the majority carrier surface mobility, μ_o, (2) zero threshold voltage, V_{TO}, and (3) the majority carrier scattering intensity in conduction channel V_{GC}.

The following five parameters characterize the physical structure of an MOS decoder operating with an output signal amplitude of 10 volts: (1) the transistor channel oxide thickness, T_{ox}, (2) the channel length, L, (3) the channel width, W, (4) the load conductance at 10 volts, G_L, and (5) the substrate resistivity, ρ.

Four additional parameters, specifying the external operating voltages, completely determine the decoder design: (1) the OFF gate voltage supply V_{OFF}, (2) the ON gate voltage supply, V_{ON}, (3) the substrate bias supply, V_{SUB}, and (4) the decoder input voltage during ON period, V_{IN}.

Measurements with p-type MOS transistors have resulted in values of μ_o and V_{GC} very close to the corresponding values for n-type transistors. The p-type transistors exhibit a negative zero threshold voltage of a few volts, whereas the n-type units usually display very little if any zero threshold voltage. To be turned ON, a p-type transistor requires a gate-to-source voltage more negative than the threshold voltage. The physical constants used for the n- and p-type transistors in the present section are listed in Table II.

The channel oxide thickness, T_{ox}, is fixed at 1000 Å. This thickness represents a compromise between reliability and transistor transconductance. The resistivity of the selected silicon substrate material is 10 ohm-cm, the value most commonly employed in the fabrication of n-type MOS transistors. An OFF gate voltage of 3 volts beyond zero pinch-off voltage is sufficient to hold an MOS transistor in a state of minimum conductance.

Table II—Physical Constants Used for n- and p-Type MOS Transistors

Type	μ_o (cm^2/V-sec)	$V_{CG}(T_{ox} = 1000$ Å) (volts)	V_{TO} (volts)
n-Type	357	15.2	0
p-Type	357	15.2	$-3.$

In order to obtain a meaningful comparison between the two decoder systems, the value of $|V_{GS}|_{MAX}$ and $|V_{SSUB}|_{MIN}$ will be standardized for the n- and p-type selection trees. The minimum value of $|V_{SSUB}|$ will be set at 4.0 volts. The values of $|V_{GS}|_{MAX}$ and $|V_{SSUB}|_{MIN}$ occur at the beginning of the turn-on period for an n-type selection tree and at the end of the turn-ON period for a p-type selection tree. The required ON gate voltages and substrate voltages are included in Table III for the different polarity decoding systems.

The input voltage, V_{IN}, becomes a dependent parameter if $|V_{GS}|_{MAX}$, W, L, G_L, and the desired 10-volt output signal are specified. These four parameters will be treated as independent variables to determine their influence on the uniformity and switching speeds of the decoder output signal.

Very little quantitative information is presently available that relates reliability and yield factors to the parameters of an MOS transistor. Previous circuit experience with MOS transistors suggests that both reliability and field are noticeably degraded under any of the following conditions:

$$L < 0.25 \text{ mil,}$$
$$W < 40.0 \text{ mils,}$$
$$T_{ox} < 1000 \text{ Å,}$$
$$V_{DS} > 20 \text{ volts (OFF state),}$$
$$V_{GS} > 34 \text{ volts.}$$

Decoders containing any transistor parameters in this range will not be analyzed, as their reliability and yield factors are considered unacceptable.

Table III—Required Operating Voltages

Type of Decoding System	$(V_G)_{ON}$	V_{SUB}
n-Type	$|V_{GS}|_{MAX}$	$-|V_{SSUB}|_{MIN}$
p-Type	$V_{IN} - |V_{GS}|_{MAX}$	$V_{IN} + |V_{SSUB}|_{MIN}$

A p-type decoder possesses a potential advantage over an n-type decoder in the gate-breakdown reliability area. One quarter (using the quaternary code) of the transistors in an n-type tree start off with maximum $|V_{GS}|$ voltage, and all but the five in the selected signal path for a 1024-output decoder remain at maximum $|V_{GS}|$ during the period of the applied input code. In comparison, the 25% of the transistors in the p-type tree that are addressed begin with minimum $|V_{GS}|$ voltage, and only those five in the selected path proceed to a state of maximum $|V_{GS}|$. Therefore the gate field breakdown stress (number of transistors with maximum $|V_{GS}|$ times duration of maximum $|V_{GS}|$) is at least two orders of magnitude greater for the n-type selection tree than for the p-type selection tree.

The output-signal uniformity is established by the values and tolerances of conductances of the selection tree and load transistors. If the ratio of the ON signal-path conductance in the selection tree to the load conductance at 10 volts is known, then the output-voltage uniformity is a unique function of the applied input voltage. The output-voltage uniformity can be measured by a voltage tolerance parameter, P, such that

$$V_{OUT} = (1 \pm P) V_o, \tag{6}$$

where V_o is the desired output-signal amplitude.

Calculation of Switching Speed

In the switching-speed calculation, it is assumed that the application of the V_{ON} voltage to the selected input code lines has preceded the initiation of an ideal voltage step at the decoder input. The output signal fall times are referenced to the instantaneous grounding of the decoder input terminal. The detrimental effect of stray intra-connection capacitances will be neglected since these capacitances represent an unknown quantity and are subject to reduction as the intra-connection technology advances. In the calculation, it is assumed that the output lines are connected to an external load capacitance of 14 pF.

The switching speed of a decoder was computed in terms of the 0-80% rise time (TR80) and the 100-20% fall time (TF20). The computer solution of differential equations of the form of Equation (1) was accurate to better than 1.0 nsec. The computed 0-90% rise times are approximately 35-40% longer than 0-80% rise times; the 100-10% fall times generally require 40-50% more time than the 100-20% fall times. The switching speeds and worst-case output-signal variations are presented in Figure 6 for selected combinations

of $|V_{GS}|_{MAX}$, G_L, W, and L, in n-type and p-type decoding systems. The worst-case output-signal variations were computed assuming 50% variations in the majority-carrier mobility within the decoder transistors. Figure 6 clearly shows that a decoding circuit using a p-type selection tree is capable of better output-signal uniformity and faster rise-time switching speed than a comparable decoding circuit with an n-type selection tree. For an identical set of $|V_{GS}|_{MAX}$, G_L, W, and L parameters, a p-type decoder typically possesses a TR80 that is 7-10% faster than an n-type decoder, a TF20 that is 30-40% slower than an n-type decoder, and a worst-case output signal tolerance that is at least
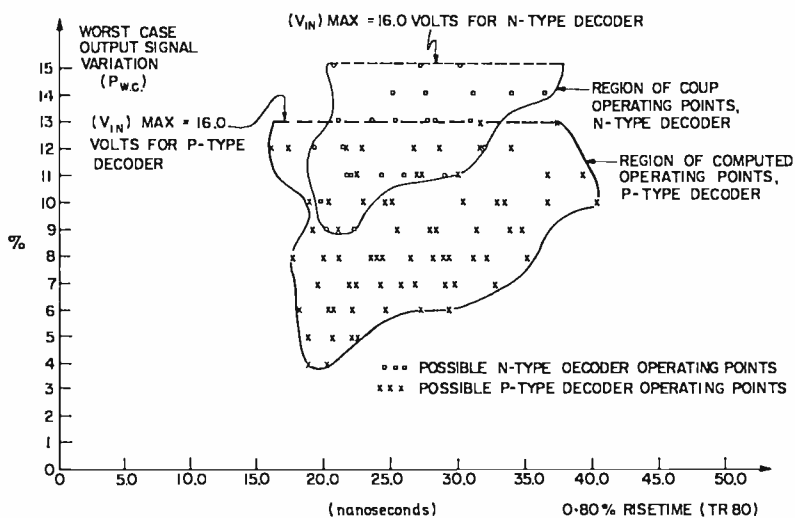
Fig. 6—Possible operating points of n- and p-type decoders in terms of output-signal uniformity and rise-time speed.

50% better than that for an n-type decoder. This conclusion results from the polarity of the output signal. If a negative-voltage output signal were required, the performance roles of the n-type and p-type selection trees would be interchanged.

The consequences of increasing the channel width of the transistors near the input end of the selection tree are examined in Table IV. The switching-speed performance of a decoder using 20, 25, 30, 35, and 40 mil channel widths in preceding levels from the decoder output terminals is compared to the performance obtained with a similar type decoder using a uniform 30-mil channel width throughout the selection tree. An increase of 10-15% is obtained in switching speed if the graduated channel widths are substituted for the uniform channel

width in the selection tree. This gain in switching speed is an increasing function of the size of the load capacitance relative to the level capacitances in the selection tree. In the present case, the level capacitances are the same order of magnitude as the load capacitances.

Table IV—Computed Switching Speed Performance of MOS Decoders with Varying Channel Widths

	All-n-Type Decoder		All-p-Type Decoder	
	TR80 (nsec)	TF20 (nsec)	TR80 (nsec)	TF20 (nsec)
"Standard Conditions" $W = 30.0$ mils $L = 0.25$ mils $\|V_{GS}\|_{MAX} = 30.0$ volts	23.6	21.9	21.1	29.4
Varying Channel Width $W_1 = 40$ mils $W_2 = 35$ $W_3 = 30$ $W_4 = 25$ $W_5 = 20$	20.1	18.8	18.2	25.1

Selection of Design Parameters

Two important conclusions are evident from the computational results. First, a 1024-output MOS decoder employing an internal quaternary code structure is capable of 0-80% rise times of less than 50 nsec. Second, there is no sharply defined set of optimum decoder parameters. Both the rise-time switching speed and the output-signal uniformity are continuously improving functions of $|V_{GS}|_{MAX}$, W, and $1/L$. Therefore, the W and $1/L$ dimensions should be maximized consistent with current technology capability and yield factors. For positive output signal applications, a p-type selection tree is preferable to an n-type selection tree in terms of turn-ON switching speed and output-signal uniformity.

The specific vehicle chosen for evaluation of the selection tree is a variation of the complementary-type decoder proposed by Burns and Gibson.[8] In the present decoder, an n-channel transistor is connected to each tree output. An output signal is initiated by simultaneously switching the n-channel transistors to an OFF state and the strobe transistor to an ON state. The necessary selection tree connections and driving waveforms for the complementary decoder are indicated

[8] J. R. Burns, J. J. Gibson, A. Harel, K. C. Hu, and R. A. Powlus, "Integrated Memory Using Complementary Field-Effect Transistors," *Digest of Technical Papers*, International Solid State Circuits Conference, p. 118, Feb. 1966.

in Figure 1. The n-channel transistors were mounted in TO-5 header cans and externally connected to the silicon slice containing the selection tree.

SELECTION-TREE FABRICATION

Topological Features

A completed selection tree on a silicon substrate and packaged in a beam-lead structure is shown in Figure 7. The selection tree requires a silicon area of 740×150 square mils. Each tree exhibits a two-column structure. The 64 outputs are available from the second col-
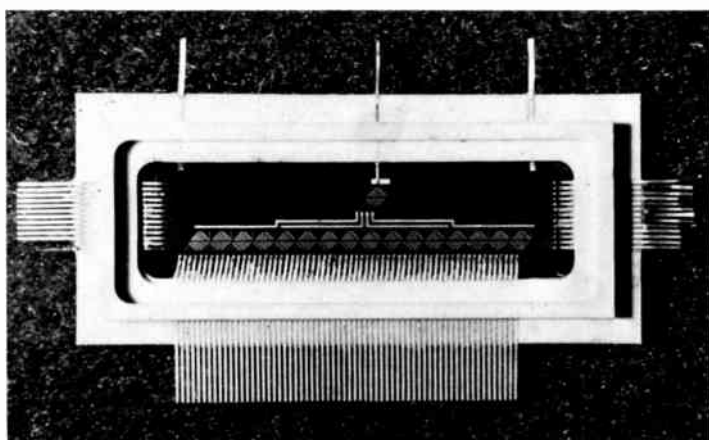
Fig. 7—Completed selection tree.

umn. The lines running the length of the columns consist of the 12-gate address lines and the signal line for the strobe transistor at the beginning of the selection tree. These 13 lines are consecutively labeled S, A_1, A_2, A_3, A_4, B_1, B_2, B_3, B_4, C_1, C_2, C_3, C_4, with C_4 being the address line closest to the output side of the selection tree. Contact tabs are provided at both ends of the address and strobe lines to facilitate the stacking of selection trees in larger selection-tree networks.

A basic group in the selection tree consists of 11 diffused (source/drain) regions. The dimensions of the diffused regions are 20×2.5 square mils and the separation between regions is 0.25 mil. The physical construction and circuit representation of such a group is indicated in Figure 8. Each MOS transistor in the group has a channel area of
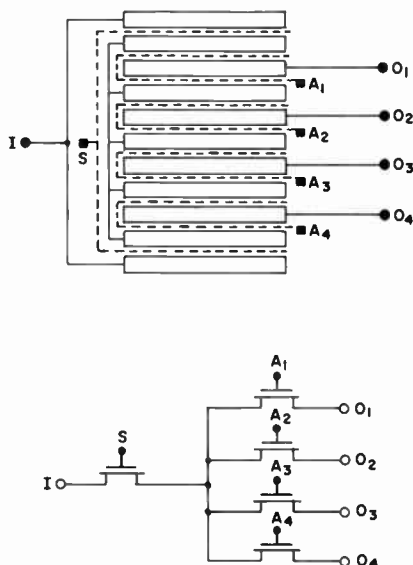
Fig. 8—Physical construction and circuit representation of a basic group.

40×0.25 square mils. The 64-output tree requires one basic group in the first column and 16 basic groups in the second column. Figure 9 is a photomicrograph showing part of the selection tree. This figure illustrates the selective oxide etching used to enable each B address line to contact every fourth group.

Fig. 9—Photomicrograph showing partial area of selection tree.

Fabrication Procedure

A flow chart of the procedures used to fabricate a selection tree is presented in Figure 10. Mechanically polished silicon wafers oriented in the [100] plane result in higher transconductance (g_m) transistors than comparable wafers that are chemically polished or oriented in the [111] plane. The starting $n+$ wafer (0.01 ohm-cm) is doped to produce a 2 to 15 ohm-cm n surface layer. The depth of the n epitaxial layer is approximately 0.6 mil. The $n+$ region provides a low-impedance contact to the evaporated metal substrate connection on the bottom of the wafer.

The basic group of transistors is constructed within an oxide well or "bathtub" area. The bathtub region is surrounded by 10,000 Å of oxide grown in a steam atmosphere. This oxide is used to insulate the intraconnection wiring from the substrate. Within the bathtub area, the 11 source/drain islands are defined by standard photoresist techniques and diffused using a boron-nitride source. A stained angle-lap view of the resulting $p+$ diffused regions is shown in Figure 11. The measured diffusion depth is 0.022 mil. There is negligible undercutting of the diffused regions into the channel area when the diffusion is performed at 975°C for 40 minutes. The resistivity of the diffused regions is in the range of 30-40 ohms/square.

The channel oxide consists of a 250 Å thermal oxide layer grown in a dry O_2 atmosphere followed by a 750 Å layer of deposited oxide doped with phosphorus. The doped oxide is commonly used to increase the electrical stability of the MOS transistor by inhibiting the mobility of charge carriers in the channel oxide.[9,10]

The most critical steps in the fabrication procedure are the etching of oxide "windows" for the source and drain contacts and the selection and application of the device metallization and crossover insulation. Failures within the finished selection tree can usually be traced to problems in one of these areas.

A 0.5-mil-wide metal gate is used to completely cover the 0.25-mil-wide channel. Any oxide pinholes in the overlap area between the metal gate and source or drain islands will result in a gate-to-source or gate-to-drain short. Oxide pinholes are most frequently introduced during the etching of the windows in the oxide prior to the source and drain metal evaporation. Clean photoresist masks and careful handling to avoid dirt contamination are essential in this operation.

[9] M. Yamin, "Observations on Phosphorus Stabilized SiO_2 Films," *IEEE Trans. Elec. Dev.*, Vol. 13, p. 256, 1966.
[10] D. Kerr, et al, "Stabilization of SiO_2 Passivation Layers with P_2O_5," *IBM Jour. Res. and Dev.*, Vol. 8, p. 376, 1964.

STARTING MATERIAL

Mechanically Polished Silicon
(100) crystal orientation

N/N⁺ epitaxial substrate

N ≃ 2-15 ohm·cm

Wafer thickness ≃ .008"

Wafer diameter ≃ 1.25"

BATHTUB AREA INSULATION

Clean Wafer

Grow 5000 Å steam oxide

Photoresist, mask (1),
etch bathtub areas

SOURCE DRAIN DIFFUSION

Clean Wafer

Grow 5000 Å steam oxide

Photoresist, mask (2),
etch source drain regions

Clean Wafer

Boron-Nitride Diffusion
975 C, 30 min.

Photoresist, mask (1),
etch bathtub areas

CHANNEL OXIDE GROWTH

Clean Wafer

Grow 200 Å dry oxide

Deposit 700 Å phosphorous
doped oxide

Grow 100 Å dry oxide

DEVICE METALIZATION

Photoresist, mask (3),
etch source drain
contact openings

Clean Wafer

Evaporate 3000 Å Al metali-
zation

Photoresist, mask (4),
etch undesired metal

GATE LINES INSULATION

Clean Wafer

Deposit 2500 Å low tem. SiO₂

Photoresist, mask (5),
etch gate lines contact
openings

Repeat with different print
of mask (5)

GATE LINES METALIZATION

Clean Wafer

Deposit { 100 Å chrome
         { 4000 Å silver
         { 100 Å chrome

Photoresist, mask (6),
etch undesired metal

Clean Wafer

Photoresist, mask (7),
etch away oxide
over input output tabs.

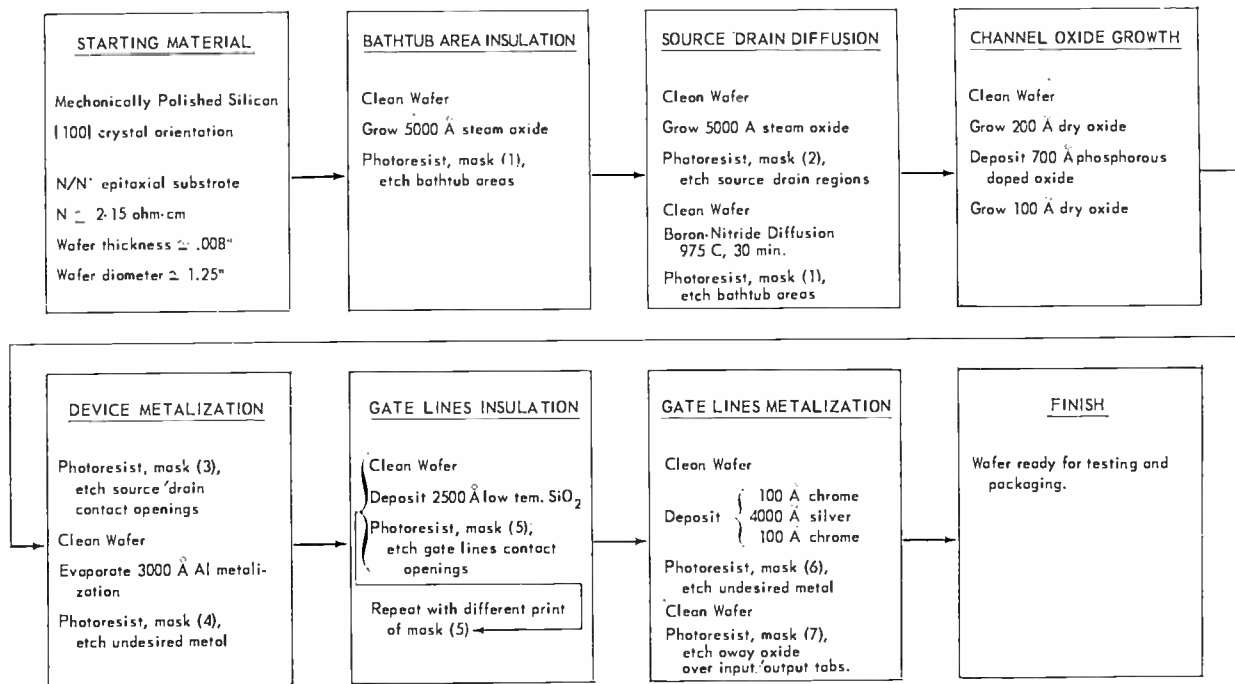**FINISH**

Wafer ready for testing and
packaging.

Fig. 10—The selection-tree fabrication process.

The limiting factor in the present work appears to be the quality of the applied photoresist—a mixture of Kodak KPR 89% and Kodak KPL 11%. The photoresist was spun dry on the wafer at 3000 rpm. The pinhole count was reduced by 60% by coating the wafer twice with photoresist before pattern exposure. A further improvement in the pinhole count was observed when the photoresist was applied directly to the wafer through a filter-loaded hypodermic syringe. The above techniques produced an average of three gate-to-source/drain shorts among the 85 transistors in a selection tree.
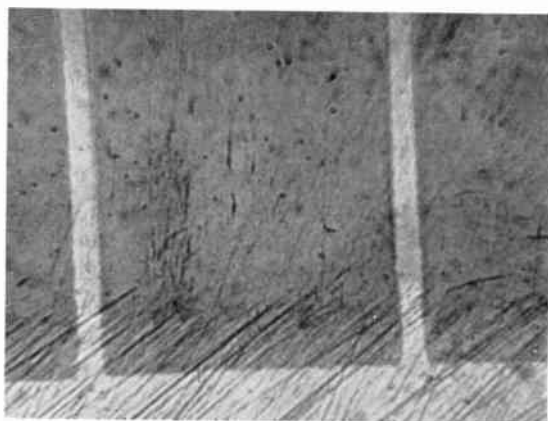
Fig. 11—Angle lap, 1000 × photomicrograph showing source/drain diffusion region.

The device (lower level) metallization and crossover insulation are interdependent and must be considered jointly. The metal and insulator materials should be chemically and electrically inert after the initial metal–insulator adherence. The metal must provide a low-impedance contact to the source/drain regions. Gate–line intraconnections require that contact openings be etched in the insulator without disturbing the underlying metal.

The insulator selected was a low-temperature-deposited SiO_2 layer. The chemical reaction involves the decomposition of silane:[11]

$$SiH_4 + O_2 \rightarrow SiO_2 + 2H_2 . \tag{7}$$

The temperature of the silicon wafer was maintained at 325°C during

---

[11] D. Flatley, *Dielectric Defects in Deposited Silicon Dioxide*, Master's Thesis, Newark College of Engineering, June 1967.
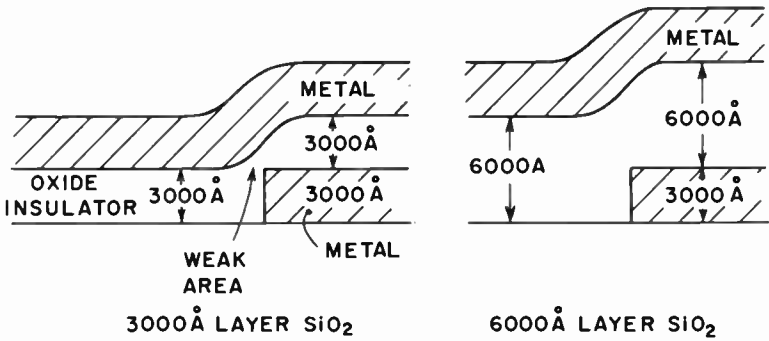
Fig. 12—Different thicknesses of deposited SiO_2.

the insulator deposition. The SiO_2 insulator is relatively simple to deposit and can be etched with good definition without disturbing the metallization systems investigated. The incidence of crossover shorts in the deposited insulator decreases as the ratio of metallization layer thickness to insulator thickness decreases. The basic problem is indicated schematically in Figure 12. Several burned crossover shorts are shown in Figure 13. The shorts usually occur along an edge of the bottom metallization pattern, the "weak" area indicated in Figure 12. The high electric field existing at conductor edges further aggravates

Fig. 13—Burned crossover shorts (within circles).

the breakdown problem. No significant decrease in crossover shorts was observed for metallization/insulator thickness ratios less than 0.6.

Two metallization procedures were studied—an all-Al evaporation and a Cr-Ag-Cr mixed evaporation. Pairs of similarly processed te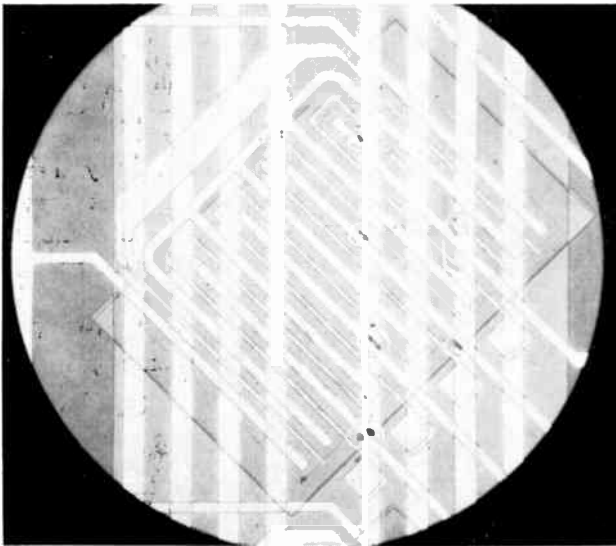st wafers were prepared to evaluate the two metallizations. Performance was measured in terms of yield and breakdown voltage data between lower and upper metallization crossover points. Using an insulator thickness of 5000 Å, a lower metallization thickness of 3000 Å, and a 200-volt test voltage, the failure rate on a Cr-Ag-Cr metallization wafer was 0.9%. On a comparable Al metallization wafer the failure rate was 0.2%. Breakdown voltages on the Cr-Ag-Cr wafer were between 250-300 volts; on the Al wafer, they were between 450-500 volts. These tests indicated the superiority of an Al metallization for the selection tree. The difficulty in the Cr-Ag-Cr metallization may be due to metal migration into the insulating oxide or poor adherence of the oxide to the silver component in the Cr-Ag-Cr structure.[12] The deposited silicon dioxide field breakdown strength was $0.8 - 1.0 \times 10^7$ V/cm.

Aluminum metallization frequently results in poor source/drain contacts, which appear as reverse bending of the I-V characteristics as seen in Figure 14(A). The solution to this problem required the Al evaporation to be done in an ultra-clean vacuum system, preferably a vac-ion system with electron-gun evaporation. Particular care was taken to ensure complete removal of the SiO_2 in the contact areas prior to evaporation.

The SiO_2 insulator was deposited and etched (for the gate–address-line contact openings) in two layers to decrease the probability of etching pinholes. Each layer is about 2500 Å. Two different photomasks, both printed from the same master mask, are used to cancel out pinholes due to acquired contact dirt. Crossover failures approaching 0.1% were obtained for the 1634 crossovers per selection tree.

Because of the large number and close spacing (10-mil centers) of input/output tabs, the selection tree required packaging in an external beam-lead structure, as shown in Figure 7, to facilitate external connections. A parallel gap welder was used to connect the package leads to the selection tree input/output tabs. The metal on these tabs must be solderable material such as silver (or Cr-Ag-Cr with the last layer of Cr less than 100 Å) or nickel to be compatible with the paral-

[12] E. Ross and J. T. Wallmark, "Increasing Life of Ag-Cr Metallization on Silicon," *Digest of Technical Papers, Reliability Physics, Sixth Annual Sym.*, Nov. 1967.

lel-gap welder. Therefore, the aluminum contact tabs on the selection tree are usually reinforced with Cr-Ag-Cr by evaporating through a metal mask or by evaporating over the entire wafer and using selective etching techniques, or with Ni by selective electroplating.

Transistor I-V Characteristics

The selection-tree transistor characteristics were recorded before application of the intraconnection wiring. The drain current was

VERTICAL SCALE: 0.5ma/DIV.
HORIZONTAL SCALE: 0.5V/DIV.

VERTICAL SCALE: 1.0ma/DIV.
HORIZONTAL SCALE 0.5V/DIV.

$V_{GS} = -12 V$
-10
-8
-6
-4
-2

$V_{GS} = -12 V$
-10
-8
-6
-4
-2

(A)
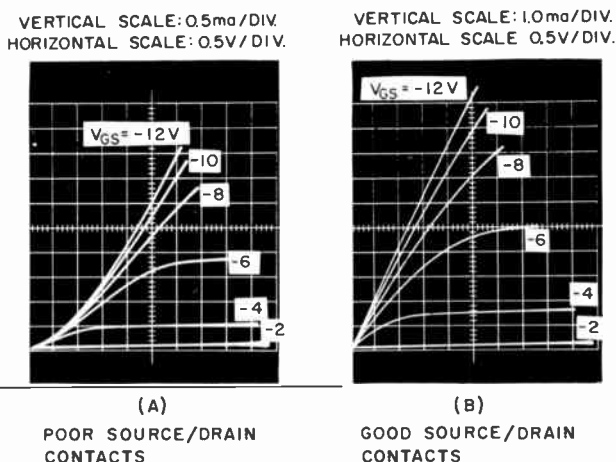POOR SOURCE/DRAIN
CONTACTS

(B)
GOOD SOURCE/DRAIN
CONTACTS

Fig. 14—Effect of poor source/drain contacts on transistor characteristics.

measured as a function of drain voltage and gate voltage. A typical characteristic for a series transistor pair is shown in Figure 14(B). The drain current and threshold voltage were then studied as a function of position on the wafer at fixed bias values. The current and voltage change gradually over the wafer surface, corresponding to localized material conditions. Overall uniformity is within 20%. There appear to be smoothly varying regions of drain-current and threshold-voltage levels.

OPERATION OF A 64-OUTPUT COMPLEMENTARY DECODER

Test System

The 64-output selection tree was connected to n-type load-switch transistors mounted in TO-5 header cans to form a complementary decoder (see Figure 1). Connections were made to the selection tree
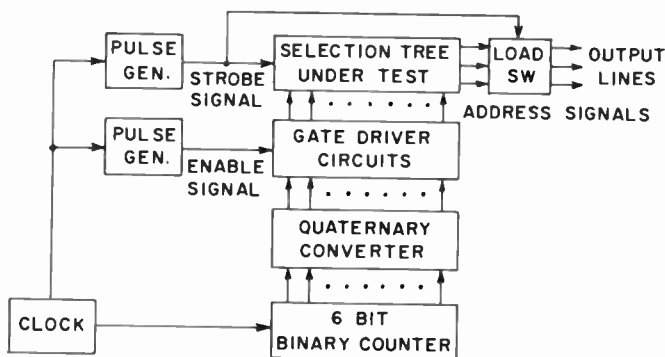
Fig. 15—Sequential addressing system.

with a 13-point in-line probe for the gate address and strobe lines and a 16-point in-line probe to test 16 consecutive output lines. The MOS load-switch transistors were characterized by a channel area of 20 × 0.4 square mils and a gate oxide thickness of 1000 Å.

A sequential addressing system was assembled as shown in Figure 15. This system tests an output line under all possible combinations of input address signals every 64 clock cycles. A good output line will produce an output signal once during 64 consecutive clock signals as shown in Figure 16.

Selection Tree Yield

The most serious problem affecting the selection-tree yield is the presence of shorts or current leakages. The observed shorts are of three general types: gate to source or drain; source to drain; and sub-strate to source or drain. The first of these types is the most frequent and the most troublesome. The failure factors of several wafers are listed in Table V.
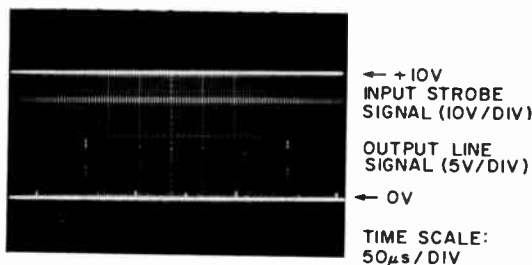

Fig. 16—Normal output-line response.

The gate-to-source or gate-to-drain shorts can result either from leakage in the thermal oxide between the metal gate and the doped source or drain regions, or from leakage between the gate and source/drain intraconnection lines through the deposited oxide insulator. Experimentally, it is difficult to distinguish whether a gate-to-source or gate-to-drain short in a finished selection tree is due to a failure in the thermal oxide or the deposited oxide. Both possibilities are electrically in parallel in the selection tree. Separate measurements of the thermal oxide and the deposited oxide indicate that failures

Table V—Selection-Tree Defects

Selection Tree	Defects					
	Shorts			Open Lines	Operable Lines	Yield
	g-s,d	s-d	subs-s,d			
J16-2	2	1	1	1	47	73%
J18-4	10	0	0	1	24	38%
J20-2	9	0	0	0	25	39%
J21-1	1	1	0	1	61	95%
J21-2	8	0	0	0	23	36%
J21-3	4	0	0	1	51	80%
J21-4	6	1	0	1	37	58%
J22-1	2	6	0	0	44	69%
J22-2	1	4	0	1	28	44%
J22-4	1	1	0	0	44	69%

within both types of oxide contribute equally to the problem. The gate-short failure rate for the 85 unconnected selection-tree transistors was 3.0 gate shorts per tree. A simulation of the selection-tree two-layer wiring with the 1634 crossovers produced a failure rate of 3.3 gate shorts per tree.

The presence of a gate short can be detected during a 64-address cycle by an output line response as shown in Figure 17(A). This figure illustrates the effect of a gate-to-output-line short in a C (third) level transistor. Because of the sequencing pattern of the quaternary input code, a C level gate is ON every fourth clock pulse (low-voltage level) and OFF during the intervening three clock pulses (high-voltage level). When the load switch is opened, the output will follow either the high- or low-voltage gate signal. When the load switch is closed and the gate signal is OFF, an intermediate voltage will appear on the output line as determined by the voltage divider network of the gate-short impedance and the closed load-switch impedance.

Source to drain shorts in the selection tree are almost invariably
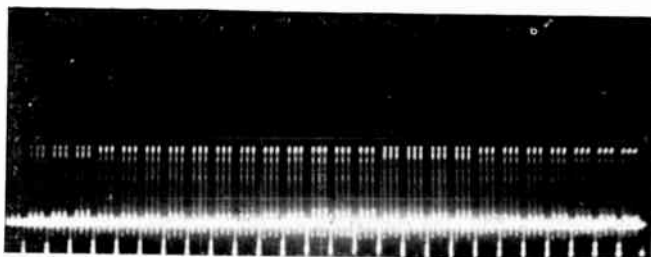
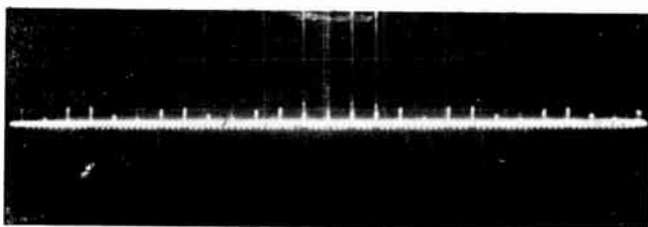Fig. 17(a)—Output-line response with gate to source or drain short in "C" level.

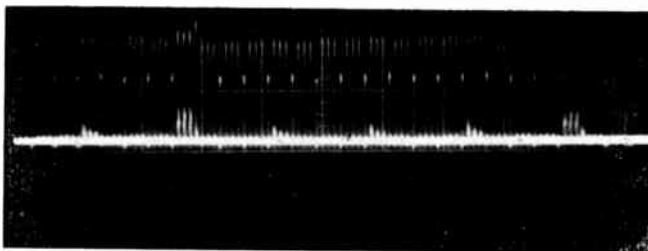Fig. 17(b)—Output line response with source to drain leakage in "B" level.

Fig. 17(c)—Output line response with substrate to source or drain short in "C" level.

due to the imperfections in the source/drain photomask. This problem is not serious in the present work and can be eliminated by using high-quality photomasks. In a wafer containing 680 channel regions (between adjacent source–drain areas) each with dimensions of 20 × 0.25 mils, there was only one occasional channel short because of a mask defect. An example of a source-to-drain short in a B level transistor is shown in Figure 17(B). The failure of the B level transistor to turn OFF introduces three extra output pulses during each 64-word cycle.

Source or drain to substrate shorts occurred even less frequently than source-to-drain shorts. The signals appearing on an output line that is shorted to the substrate is shown in Figure 17(C). The display is similar to that obtained from a gate short, except the substrate is always at a positive voltage instead of returning to a low (ON) voltage every fourth cycle.

Discrimination Ratio

The address sequencing program also permitted a check of cross talk between output lines. A measure of this cross talk is the discrimination ratio—the ratio of the maximum signal on an unselected output line to the minimum signal appearing on a selected line. The output signal on an unselected line may occur from current leakage through one or more OFF MOS transistors. The magnitude of this output signal is proportional to the strobe pulse duration and inversely proportional to the load capacitance on the output side of the MOS transistor(s). Barring catastrophic transistor failures, the discrimination ratio was less than 0.05 for output lines operating with a load capacitance of 80 pF and a strobe pulse width of 300 nsec.

If the load capacitance is known, an estimation of the minimum MOS OFF resistance can be obtained by simple RC network analysis. Under the above operating conditions, the minimum OFF resistance for a single selection-tree MOS transistor must be >100,000 ohms.

Switching Speed

The switching speed of a decoder incorporating the selection tree depends on such factors as the applied bias voltage and driving waveforms, the characteristics of the MOS selection-tree transistors, and the loading conditions on the output line. For a decoder with an output signal amplitude of +10 volts, the following bias voltages are used in the circuit of Figure 1:

$$V_{\text{IN}} = +10,$$
$$V_{\text{OFF}} = +10,$$
$$V_{\text{ON}} = -4,$$
$$V_{\text{SUB}} = +15.$$

In all cases, the V_{OFF} voltage on the gate address lines is sufficient to hold the selection-tree transistors in an OFF state, since the threshold voltage, V_T, is always greater than 2 volts (source-to-gate voltage) under all operating conditions. The strobe signal is normally at $+10$ volts (OFF) and decreased to 0 to turn on the decoder. The 90% to 10% fall time of this strobe signal is 20 nsec. Figure 18(A) shows a timing diagram for the decoder operating at a 2 MHz rate. Figure

(a)

(b)
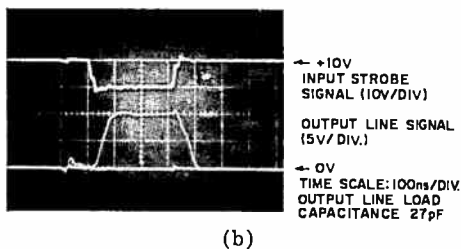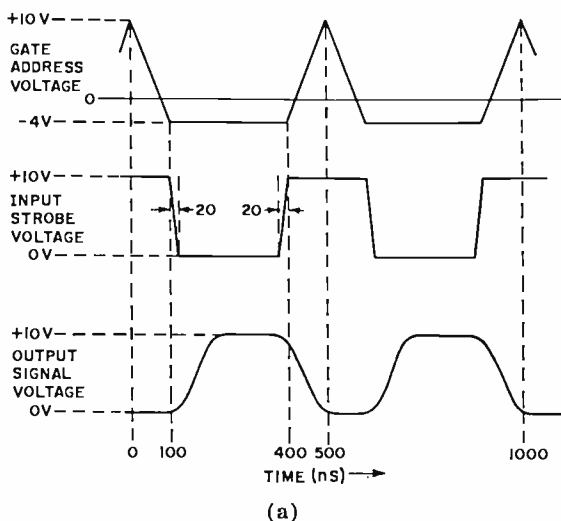
Fig. 18—(a) Decoder timing diagram and (b) decoder output signal.

18(B) illustrates an actual output signal on and output line with a load capacitance of 27 pF.

The switching speeds of decoders using selection trees of several different wafers are compared in Table VI for a load capacitance of 88 pF. The selection-tree starting material and electrical characteristics of four transistors connected in series are also included in this table. The fastest switching speeds are associated with the high transconductance (g_m) and low source and drain to substrate capacitance

Table VI—Selection-Tree Characteristics and Switching Time

	Starting Material			Electrical Characteristics Four Series Transistors			Switching Times (nsec) $C_L = 88$ pF		
Wafer	Orien-tation	Surface Doping	Surface Resistivity (ohm-cm)	$V_{T(g-s)}$ ($V_{SSUB} = 0$) (volts)	I_D $\|V_{DS}\| = 2.5$ V $\|V_{GS}\| = +10$ V $V_{SSUB} = 0$ V (mA)	g_m $\|V_{DS}\| = 2.5$ V $\|V_{GS} - V_T\| = 6$ V (μmhos)	Rise Time)—10% 10—90%		Fall Time 90—10%
J16	[111]	n	2	−4	3.5	480	61	188	83
J17	[100]	n	2	−3	4.0	475	60	150	110
J18	[100]	n/n+	15	−2	3.3	330	40	98	80
J21	[100]	n/n+	15	−2	6.3	750	40	70	90
J22	[100]	n/n+	2	−3.5	6.1	750	63	132	78

wafers. The high-transconductance wafers are usually oriented in the [100] direction; the low-capacitance wafers have higher values of surface resistivity (15 ohm-cm). The apparent discrepancy between the actual switching times obtained from the 64-output decoder and those computed for the 1024-output decoder arise primarily from the increased value of load capacitance (88 pF versus 14 pF). Also, the use of finite-rise-time input signals and the additional strobe level incorporated in the 64-output tree act to reduce the switching speed.
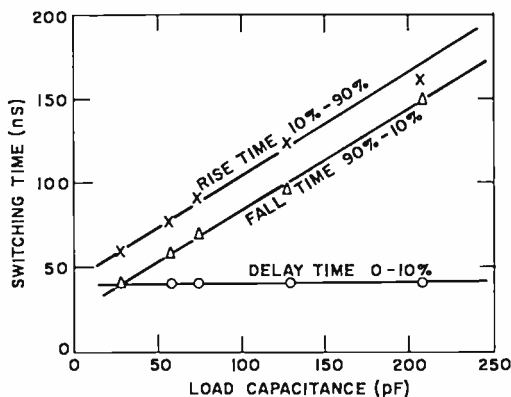
Fig. 19—Switching time as a function of load capacitance.

Switching-time data as a function of load capacitance are presented in Figure 19. The 0 to 10% delay time is determined primarily by the internal selection-tree conductances and capacitances, and is relatively independent of the load capacitance. For load capacitances greater than 27 pF, the output rise times and fall times vary linearly with load capacitance, suggesting a simple RC charging relation of the following form:

$$T_{10\to90\%} = R_{\text{equiv}}C_L \ln 9 \tag{8}$$

If the load capacitance and rise time are known, this equation can be solved for R_{equiv}, the effective resistance of the ON path in the selection tree. For a rise time of 100 nsec and a load capacitance of 90 pF, this effective ON path resistance is 500 ohms. This value is in general agreement with those obtained by direct ohmic measurement, which are in the 300-500 ohm range.

The influence of the various voltage biases on the switching times was studied. The rise time (10-90%) as a function of substrate bias,

ON gate-address voltage, and ON strobe voltage are shown in Figure 20. Increasing the substrate bias reduces both the source and drain to substrate capacitances, which should decrease the rise time, but the transistor conductances (at fixed gate voltage) are also reduced through interaction with the threshold voltage. The latter effect should
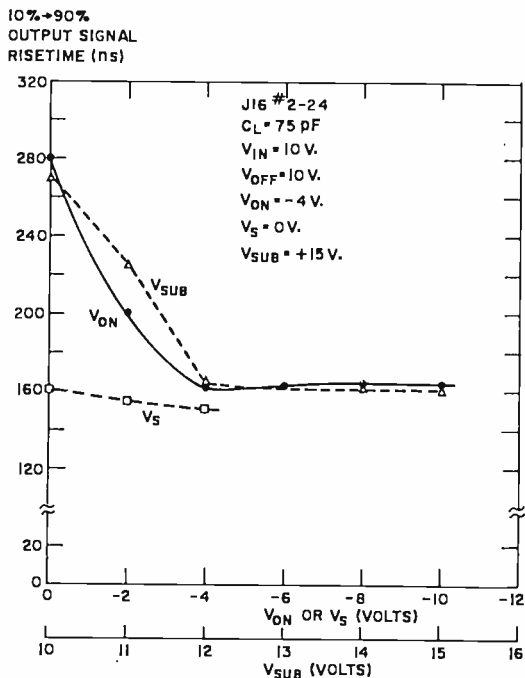
Fig. 20—Decoder rise time as a function of bias voltage.

increase output signal rise time. For increases in V_{SUB} greater than 12 volts, very little change in rise time is noted, as seen in Figure 20. Increasing the ON address-gate voltage, V_{ON}, increases the conductance of transistors within the tree (decreasing resistance of R_{equiv} which charges C_L) and therefore decreases the rise time. For V_{ON} less than −4 volts, the tree transistors approach maximum conduction and further rise time improvement is slight. Similarly, decreasing the ON strobe voltage below 0 volts has relatively little effect on the rise time.

The fall time is relatively independent of bias-voltage settings and remains at approximately 80 nsec. This fall time is determined principly by the ON resistance of the load-switch transistor and the size of the load capacitance. The 0 to 10% turn-on delay time, typically 40-60 nsec, has also been observed to be relatively independent of the bias voltages.

Input Power Requirements

Both real and reactive power must be supplied to the decoder. The real power is supplied from the V_{IN} voltage source and is dissipated through the tree source drain lines and the load switch. This power has been measured at 20 mW under operating conditions of 2 MHz and 27 pF output capacitance. For output load capacitances sufficiently greater than the internal tree capacitances ($\sim$20 pF), a relation between power dissipation and operating speed has been derived by Burns.[13]

$$P_{IN} = C_L V_{IN}^2 f_o \tag{9}$$

where C_L is the output load capacitance and f_o the frequency of operation.

The voltage swings on the transistor gate lines are responsible for the reactive power. Measured capacitance values for the different types of tree gate lines to the substrate material are:

> A address line: 52 pF,
> B address line: 64 pF,
> C address line: 90 pF,
> S strobe line: 14 pF.

For 2-MHz decoder operation, the transition times (10 to 90%, 90 to 10%) were set at 100 nsec for the address lines and 20 nsec for the strobe line. A 14-volt transition on the address line requires a current $[I = C(dv/dt)]$ of 29 mA and a reactive power of 116 mW. A 10-volt transition on the strobe line requires a current of 7 mA and a reactive power of 84 mW.

CONCLUSION

This paper has discussed the design and fabrication of a 64-output MOS selection tree and the performance of a complementary decoder circuit using the selection tree. The decoder is designed to produce output signals of +10 volts amplitude. A load capacitance of approximately 90 pF is used to simulate the loading effect of MOS transistor gates being driven by the decoder. The discrimination ratio between an unselected output line and a selected line is less than 5% for an output pulse duration of 250-300 nsec. The decoder is capable of operating in a moderate-speed (>1 MHz) digital system. Typical delay times and rise times of the decoder are 50 nsec and 100 nsec respectively. Fall times, determined primarily by the discrete load-switch

---

[13] J. R. Burns, "Switching Response of Complementary-Symmetry MOS Transistor Logic Circuits," *RCA Review*, Vol. 25, p. 646, Dec. 1964.

transistors, are approximately 80 nsec. Real power consumption at 2 MHz operation is 20 mW. Reactive power required to charge the gate address and strobe lines is approximately 200 mW. The yield of operable output lines in selection trees that have survived the fabrication process varies between 36-95%. The average yield of 10 selection trees is 60%. The primary cause of non-operable output lines is gate leakages. Three techniques to reduce the number of gate shorts within a selection tree are suggested for future work.

(1) The metallization topology can be redesigned to reduce the number of metal crossovers. By expanding the size of the selection tree and shifting the gate address lines to the areas between transistor columns, a reduction from 1,684 to 415 crossovers is possible.

(2) Semiconductor tunnels can be used at crossover points. Degenerately doped semiconductor material can be substituted for the bottom-layer metallization at crossover points. An oxide insulator should adhere better to the degenerate semiconductor than a metal because of similar crystal structure and elimination of the vertical edges around a raised metal line.

(3) Use a different insulator or a different method of applying a silicon oxide insulator can be used to lower the crossover defect rate.

The MOS selection tree discussed herein has the property of establishing a unique low-impedance path from the input to a selected output line. Possible areas of application include display and information processing and storage systems.

MONOLITHIC SENSE AMPLIFIER FOR
LAMINATED-FERRITE MEMORIES

BY

H. R. BEELITZ

RCA Laboratories
Princeton, New Jersey

Summary—A high-sensitivity (1 mV) moderate-bandwidth (10 MHz) ac-coupled sense amplifier has been designed and integrated. The amplifier, designated TA5196, includes provision for strobing, detecting, pulse forming, and internal inversion logic. It can be used with a split-sense-line-organized laminated-ferrite memory.

A design technique called "thermal feedback," which permits the fabrication of a high gain all dc-coupled integrated sense amplifier with minimum chip area, has been proven feasible. The input imbalance of a discrete-component high-gain (60 dB) d-c amplifier was controlled to within ±30 microvolts with thermal feedback. This represents an improvement by a factor of 100 over the identical amplifier using conventional design techniques.

INTRODUCTION

I N SUPPORT of the laminated-ferrite-memory program at RCA Laboratories,[1] a design study was initiated to determine the feasibility of integrating a very high sensitivity (1 mV), moderate-bandwidth ($\geqq$ 3 MHz) sense amplifier.

The broad specifications of the design objective are listed in Table I. By far the most stringent specification is the requirement of high sensitivity. The amplifier should be capable of "sensing" a 1-mV pulse and amplifying it sufficiently to drive standard logic circuitry. This requires an overall gain of 60 dB to provide a 1-volt output signal.

The specified ± 2-volt digit pulse preceding the 1-mV signal pulse poses another severe problem. The recovery of the amplifier from an overdrive condition due to this "digit blast" must be sufficiently fast for it to be able to sense the small 1-mV signal.

The specification that the amplifier be intergratable arises, of course, for reasons of economy and reliability. In addition, the monolithic environments offers certain opportunities in design (transistor matching, temperature tracking, etc.). On the other hand, area and

---

[1] R. Shahbender, C. P. Wentworth, K. Li, S. E. Hotchkiss, and J. A. Rajchman, "Laminated Ferrite Memory", *RCA Review*, Vol. 24, p. 705, Dec. 1963.

power constraints arise due to the yield problem inherent in mono-
lithic construction. Chip size should be kept as small as feasible to
maximize yield. This is especially important if arrays of sense ampli-
fiers are to be fabricated.

The differential amplifier was chosen as the basic building block
for the sense amplifier. This was done to take advantage of the high
common-mode rejection that is possible with the differential amplifier
for limiting overdrive conditions while writing into the memory. The
differential amplifier also lends itself particularly well to monolithic
fabrication, since its performance characteristics (gain, temperature
stability, etc.) are primarily dependent upon resistance ratios, device
matching, and tracking, etc., that can be readily achieved in mono-
lithic form.

Table I—Sense Amplifier Nominal Specifications

Gain (Overall)	60 dB
Bandwidth	3 MHz
Digit Pulse, Maximum	± 2 V
Threshold	1 mV
Output Voltage	1 V
Temperature Range	$10°C$ to $65°C$

Since a complete sense amplifier is proposed, provision for strob-
ing, thresholding, pulse forming, internal logic, etc must be provided.
This must be done within the constraints of minimum area (com-
ponent count) and power as noted before.

A further complication is the requirement that the amplifier be
used in conjunction with a single crossover-per-bit, split-sense-line-
organized laminated-ferrite memory. Figure 1 shows a schematic dia-
gram of the split-sense-line laminated-ferrite memory and differential-
input sense amplifier.

SPLIT-SENSE-LINE LAMINATED-FERRITE MEMORY

The motivation behind using a split-sense-line arrangement for the
laminated-ferrite memory is twofold.

1. Memory capacity for a given sense signal can be increased, per-
haps doubled, due to the reduction in sense-signal attenuation. Re-
duced attenuation is a result of the shortened signal-path length
obtained with the split line.

2. The high (several volts) digit blast back voltage present during
digiting can be coupled to both inputs of a differential input sense
amplifier. Amplifier over-drive is thereby considerably reduced due
to the high common-mode rejection of the input differential stage.

The split sense line imposes a logic problem, however, in that a "one" signal on one half of the split line produces the opposite differential amplifier (D.A.) output to that for a "one" signal on the other half of the split line. In addition, for bipolar inputs, a "one" on, say, the top split sense line produces the same D.A. outputs as that for a "zero" on the bottom sense line. Clearly then, in order to be able to
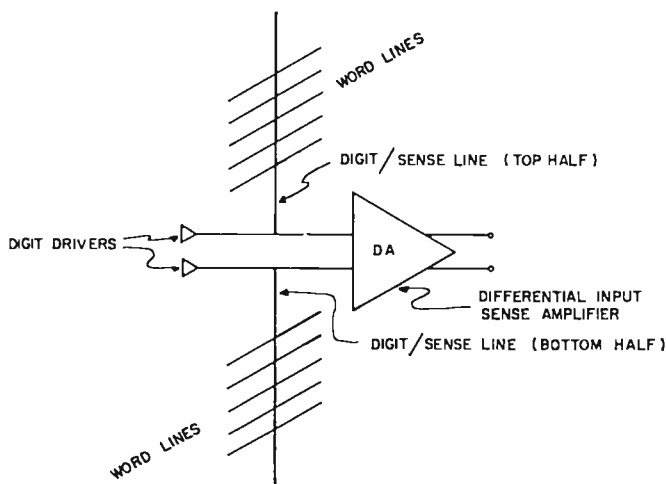
Fig. 1—Single-crossover-per-bit, split-sense-line organization of laminated-ferrite memory.

distinguish a sensed "one" from a sensed "zero", it is necessary to know in which half of the split sense line the signal originates (from the contents of the current address register) and then to perform the necessary inversion logic prior to detection.

Of course, digiting could be done so as to generate the proper sense signals (say a positive sense signal on the top line and a negative sense signal on the bottom line for a stored "one"), which would result in a single D.A. output polarity. However, the digit blast would then appear differentially across the D.A. inputs, creating severe overload problems. This approach does not appear to be very practical.

The required logic could also be performed at the output of the differential sense amplifier. This requires, for simplified gating, that the amplifier be kept differential throughout, precluding the use of "single ending" with the resultant savings in component count and chip area (important for integration) and, of course, in cost. The additional gates required at the amplifier output increase signal propagation time as well as further increasing circuit complexity and cost.

The approach adopted was to perform the required logic (inversion) internal to the sense amplifier. The circuit shown in Figure 2, which is discussed later, provides the required logic.

ALTERNATIVE APPROACHES

A number of approaches for implementing the high-gain sense amplifier were examined. These are outlined below along with a brief discussion of their applicability.
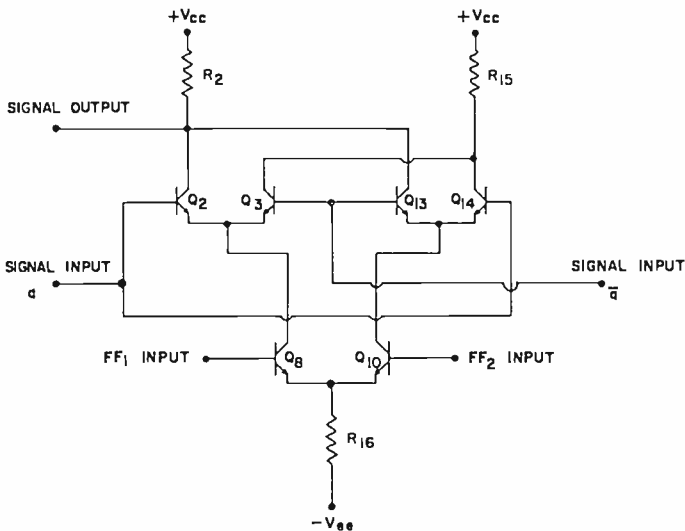
Fig. 2—Current-steered logic differential stage.

A. All-dc-Coupled Amplifier. In an all-dc-coupled differential sense amplifier, the sense signal must be larger than the inherent d-c offset uncertainty. Since the principal offset is due to the V_{bc} mismatch of the input differential transistors, typically 5 mV, it is obvious that only signals greater than 5 mV can be discriminated. A 1-mV signal cannot be sensed unless the input offset uncertainty is reduced to less than 1 mV. Since a completely monolithic structure is proposed, individual device matching is not possible. Trimming with external resistances is possible but certainly not desirable, as it is neither economic nor compatible with the monolithic approach.

B. All-dc-Coupled Amplifier with Offset Error Sensing and Correction. This approach can be further characterized by the method of error correction employed, electrical or thermal.

1. Offset Correction by Electrical Means

This method requires error detection (offset detection) during some quiescent period of the memory cycle (no input signal) and electrical feedback of a correction signal. Strobing and analog memory are required in the feedback loop. A large capacitor would probably be required for implementing the analog storage function. Since large-value capacitors are incompatible with monolithic fabrication, this technique is not practical.

2. Offset Correction by Thermal Means

Here, again, error detection during a quiescent period is required, but the feedback loop is completed via thermal coupling. This method, which we may term "thermal feedback",[2] uses the temperature dependence of the base–emitter characteristic of the input differential transistors as the control parameter, thus permitting large time constants and sensitive control without reactive elements or electrical loading or disturbance of the input stages. The scheme has the advantage (as does electrical feedback) of correcting imbalance dynamically regardless of the cause. The technique appears feasible, and experiments with integrated structures and amplifiers utilizing thermal feedback are described later.

C. AC-Coupled Amplifier. This approach is the most obvious and straightforward solution to the dc-offset problem. Unfortunately, it introduces new problems associated with the repetition rate and duty cycle of the signal pulses and the difficulty of obtaining the long time constant required during signaling and a short time constant during digiting. In addition, only low values of capacitance (less than 50 pF) are practical with the monolithic approach.

THERMAL FEEDBACK FOR AUTOMATIC OFFSET CORRECTION

The advantages that accrue from using some type of built-in automatic offset correction thermal feedback rather than the apparently simpler a-c coupling are readily seen. Ninety percent of similar transistors on a chip typically have V_{be}'s matched to within 5 mV. However, in a high-gain dc-coupled amplifier, even these relatively small imbalances would be sufficient to saturate the output stage, even in the absence of a signal, unless a sufficiently broad voltage range were allowed. This is, of course, costly in terms of excessive dissipation. Capacitive coupling from the front-end dc-coupled amplifier (comprising perhaps several stages) to the threshold detector would

---

[2] J. Amodei, private communication.

still be required in order to distinguish the amplified 1-mV signal pulse from the amplified 5-mV input dc offset. In view of the above and the previous discussion of the alternative techniques, a preliminary investigation of thermal feedback for automatic offset correction was undertaken.
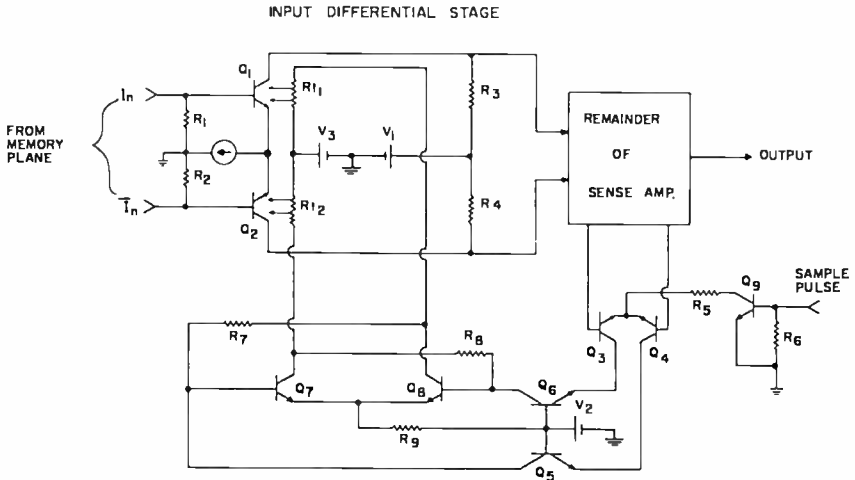
Fig. 3—Differential amplifier with thermal feedback.

EXPERIMENTS WITH THERMAL FEEDBACK

A breadboarded high-gain (60 dB) dc-amplifier test vehicle for the thermal-feedback concept was constructed using discrete components. A diagram of the test amplifier is shown in Figure 3. With the thermal-feedback loop open, the output offset (input offset multiplied by the amplifier gain) was measured to be 3 volts for a given set of devices and resistors. With the thermal-feedback loop closed, the output imbalance was controlled to within ± 30 millivolts. This represents an effective input imbalance of ± 30 microvolts, or an improvement by a factor of 100.

The flip-flop shown in the bottom part of the diagram stores the offset polarity as sensed during the quiescent period. Storage is performed so that power may be continuously dissipated in the appropriate heat-source resistor R_{t1} or R_{t2}, thus generating the required thermal difference between input transistors Q_1 and Q_2. As indicated in the diagram, R_{t1} is tightly coupled (thermally) to transistor Q_1, while R_{t2} is tightly coupled to transistor Q_2. The flip-flop is set to the

proper state during the sampling period by energizing strobe transistor Q_9.

To determine the feasibility of implementing the thermal feedback approach in a monolithic structure, a number of tests were made of thermal coupling and temperature difference between elements of avail-
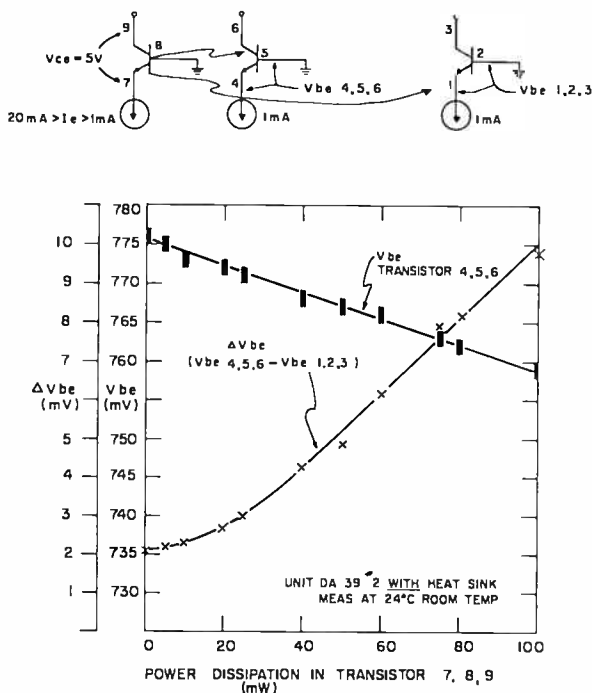
Fig. 4—Thermal coupling tests (1).

able integrated circuits. Plots of these test results are given in Figures 4 and 5. A photograph of the test structure used is shown in Figure 6.

The tests results indicate that, for the test structure used, a 2-mV differential in V_{be} can be developed with less than 50 mW dissipation. Of course, the test structure is not optimum; the heat-source transistor is too distant from its associated differential amplifier transistor. A closer spacing (more readily obtained with a heat-source resistor) would reduce the power dissipation required for a given thermal difference.

The thermal-coupling experiments discussed above are interesting,
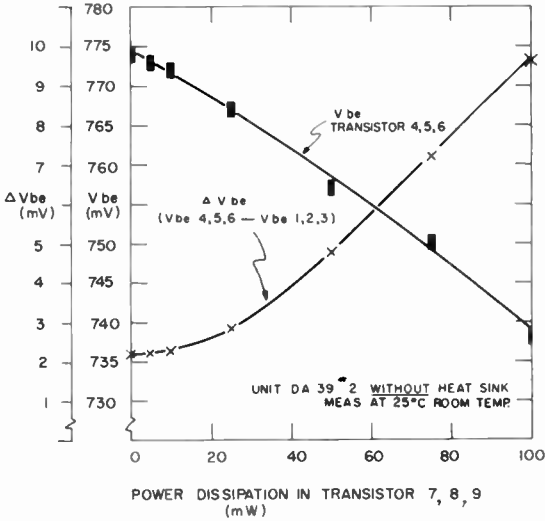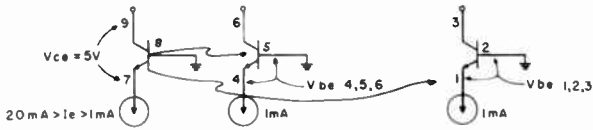
Fig. 5—Thermal coupling tests (2).
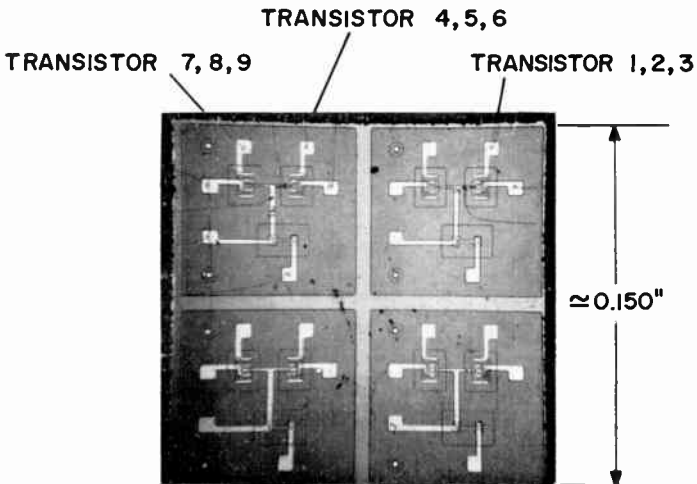
Fig. 6—Integrated structure used for thermal coupling tests.

but are not conclusive as to the practicality of incorporating thermal feedback into a monolithic amplifier. We decided, therefore, to proceed with the design and integration of the more straightforward (but marginal) ac-coupled sense amplifier. Since the differential amplifier stages of the thermal feedback amplifier are very similar to the differential amplifier stages of the ac-coupled amplifier, a large portion of the thermal-feedback sense amplifier could also be verified with the integration and testing of the ac-coupled amplifier. To complete the measurements required for designing the thermal-feedback circuit in monolithic form, two resistor–transistor pairs were included on the chip. These allow more realistic determination of the chip-dissipation–thermal-gradient parameters.

AC-Coupled Sense Amplifier

The ac-coupled sense-amplifier circuit that was developed is shown schematically in Figure 7. Essentially, the amplifier is composed of a two-stage high-gain differential amplifier capacitively coupled to a strobed detector. By single ending the output of the second stage, a relatively simple circuit requiring a low total value of coupling capacitance (22 pF) resulted. However, amplifier performance had to be kept to marginal levels in order to hold the circuit complexity and dissipation to a minimum.

The need for maintaining extremely close device matching has been partly circumvented by a one-shot correction scheme. The correction consists of shorting out, external to the flat pack in which the integrated circuit chip is mounted, ten percent of either collector resistor of the first stage, depending on the imbalance polarity. This effectively reduces the input imbalance (offset) from 5 mV to less than 3 mV. The remaining d-c imbalance is eliminated by the a-c coupling prior to detection.

The second stage of the D.A. provides additional signal gain while incorporating current-steered logic for inversion. The need for internal inversion logic was discussed previously.

The combination analog and digital circuit is shown separately in Figure 2. Note that the upper transistors (Q_2, Q_3, Q_{13}, Q_{14}) of what appears to be a logic circuit are also in the signal path and, additionally, provide analog gain of the sense signal, functioning as an integral part of the amplifier. The logic signal inputs (at ECCSL current-mode logic levels of -0.8 V ["one"] and -1.6 V ["zero"]) labeled FF_1 and FF_2 are derived from the address register. They determine whether or not sense-signal inversion will occur prior to detection.

The single-ended output of the second stage is capacitively coupled
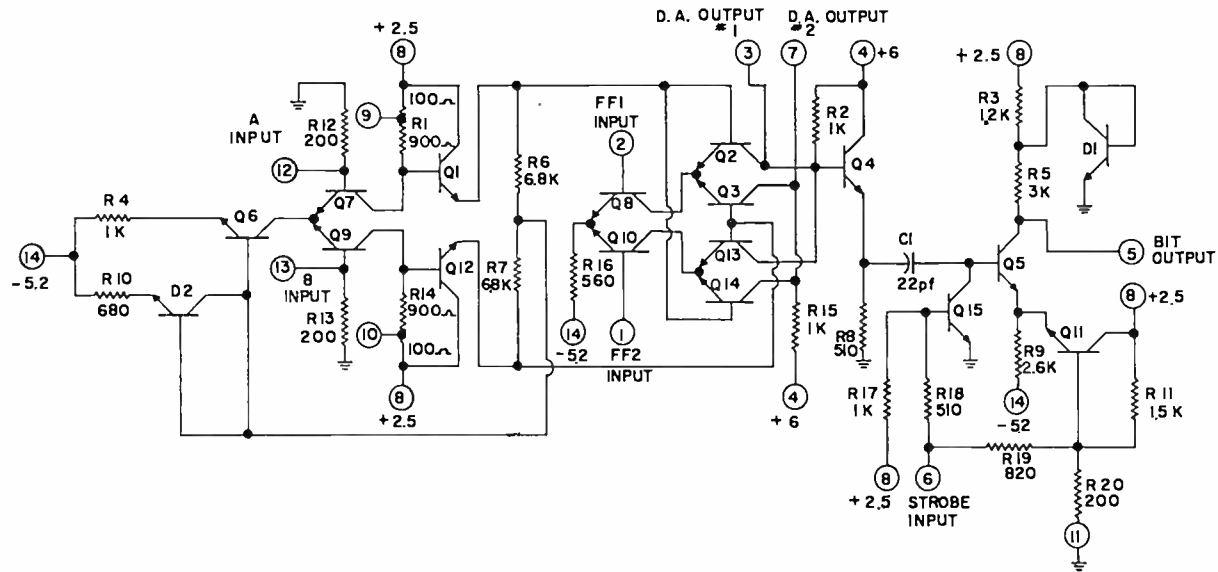
Fig. 7—Schematic of ac-coupled sense amplifier.

via a 22-pF capacitor to the base of the saturating transistor Q_5. The strobed clamp transistor Q_{15} effectively bypasses any signal to ground when saturated. When Q_{15} is cut off, a 0.5-volt positive signal at the base of Q_5 is sufficient to turn on Q_5, saturating it briefly and developing the output pulse at its collector. The emitter voltage of Q_5 (−0.45 V at 25°C), which determines the detector threshold, is set by emitter followed Q_{11}. Because of the temperature tracking of detector transistor Q_5 with the threshold setting transistor Q_{11}, the threshold is relatively independent of temperature. The amplifier output is taken from the collector of Q_5.
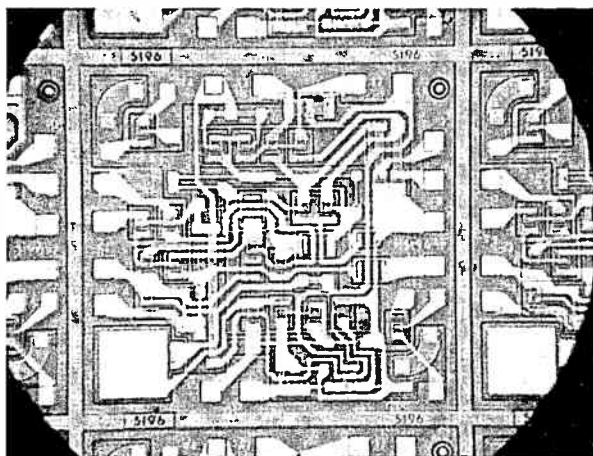
Fig. 8—Integrated sense amplifier chip (TA5196).

INTEGRATED SENSE AMPLIFIER

The ac-coupled sense amplifier circuit of Figure 7 was integrated by RCA Electronic Components, Somerville, N. J. The circuit has been given the development designation of TA5196. It has been fabricated with standard, N+ pocket, and gold-doped processes. A photograph of a complete circuit on a chip is shown in Figure 8. The chip measures 70 by 75 mils. Note that the chip also contains (at the upper left and lower right corners) the resistor–transistor pairs required for thermal testing for the thermal-feedback amplifier. They are not electrically a part of the sense amplifier. The large square area at the lower left corner is the 22-pF coupling capacitor. Little attempt was made to optimize the layout for minimum area, since the integrated chip serves primarily as a design test and verification vehicle.

SENSE AMPLIFIER EVALUATION

Operating Level and Input Offset Tests

Evaluation of the integrated sense amplifier involved both d-c (level) and a-c (pulse) testing. The d-c testing consisted essentially of setting up the flat-pack-mounted amplifier in a test jig, applying nominal power supply voltages, appropriate strobe and logic levels, and noting the circuit operating levels. Correct operation of the inversion

Table II—Standard Process Samples, DC Tests

Unit No.	Circuit Connection[1]	D.A. Normal Output[2]		D.A. Inverted Output[3]	
		#1 (volts)	#2 (volts)	#1 (volts)	#2 (volts)
1.	Normal[4]	+2.83	+1.31	+0.74	+2.82
	$9 \rightarrow 8$[5]	0.79	2.74	2.89	1.43
	$10 \rightarrow 8$[6]	4.81	3.04	0.69	2.99
2.	Normal	2.44	3.88	3.43	2.92
	$9 \rightarrow 8$	1.29	5.02	4.59	2.99
	$10 \rightarrow 8$	3.73	2.60	2.16	4.20
3.	Normal	1.92	4.55	4.25	2.71
	$9 \rightarrow 8$	0.99	5.50	5.13	3.41
	$10 \rightarrow 8$	3.28	3.16	2.99	3.48
4.	Normal	2.76	3.31	2.66	3.36
	$9 \rightarrow 8$	1.26	4.58	4.24	2.69
	$10 \rightarrow 8$	4.15	2.54	1.21	4.61
5.	Normal	2.55	3.79	3.66	2.62
	$9 \rightarrow 8$	2.06	4.28	4.13	2.60
	$10 \rightarrow 8$	3.06	3.26	3.15	3.14

[1] —see Fig. 7 for pin locations.
[2] —with amplifier logic input labeled FF_1 set to -0.8 V
[3] —with amplifier logic input labeled FF_2 set to -0.8 V
[4] —pins 9 and 10 are left open
[5] —pin 9 is shorted to pin 8
[6] —pin 10 is shorted to pin 8

and offset correction circuitry is readily checked. Tables II, III, and IV list the results of the dc-level tests made on the sample sense amplifiers.

The columns labeled "Circuit Connection" refer to the single-shot offset correction technique discussed previously. Input offset before correction can be determined from the tabulated data. The approximate input offset can be obtained by dividing the differential output offset (the difference between D.A. output No. 1 and D.A. output No. 2) by the nominal amplifier differential gain (1000). Calculated input offsets for the gold-doped amplifiers of Table IV are given in Table V.

Note that, for the sample sense amplifiers tested, the effective input offset is less than 2 mV. Since the design allowed for up to 5 mV

Table III—n+ Pocket Samples, DC Tests

Unit No.	Circuit Connection[1]	D.A. Normal Output[2]		D.A. Inverted Output[3]	
		#1 (volts)	#2 (volts)	#1 (volts)	#2 (volts)
6.	Normal[4]	+2.76	+3.21	+3.19	+2.73
	$9 \to 8$[5]	1.39	4.71	4.55	2.83
	$10 \to 8$[6]	4.25	2.59	1.70	4.32
7.	Normal	2.03	4.13	3.99	2.44
	$9 \to 8$	0.99	5.21	5.05	3.24
	$10 \to 8$	3.52	2.57	2.49	3.65
8.	Normal	2.11	2.95	3.81	2.20
	$9 \to 8$	1.01	3.85	4.94	3.06
	$10 \to 8$	3.65	2.09	2.29	2.84
9.	Normal	1.64	4.49	4.41	2.74
	$9 \to 8$	0.87	5.03	5.26	3.41
	$10 \to 8$	3.09	3.10	2.96	3.15
10.	Normal	2.01	3.84	3.64	2.25
	$9 \to 8$	0.95	4.87	4.86	3.11
	$10 \to 8$	3.60	2.26	2.09	3.77

Notes: (See Table II).

offset, this is well within tolerance. As Tables II through IV indicate, offset correction by externally shorting out 10% of either input collector resistor does not result in further improvement with these already small offsets. The offset-correction circuitry could be easily modified,

TABLE IV—Gold-Doped Samples, DC Tests

Unit No.	Circuit Connection[1]	D.A. Normal Output[2]		D.A. Inverted Output[3]	
		#1 (volts)	#2 (volts)	#1 (volts)	#2 (volts)
1.	Normal[4]	+3.11	+3.11	+2.59	+3.59
	$9 \to 8$[5]	1.32	4.91	4.63	1.52
	$10 \to 8$[6]	4.87	1.34	1.16	5.04
2.	Normal	3.23	2.92	2.91	3.22
	$9 \to 8$	1.44	4.79	4.69	1.41
	$10 \to 8$	4.82	1.31	1.29	4.91
3.	Normal	4.10	2.05	2.07	4.05
	$9 \to 8$	1.97	4.17	4.15	1.95
	$10 \to 8$	4.97	1.25	1.19	5.01
4.	Normal	3.36	2.62	1.86	4.11
	$9 \to 8$	1.40	4.60	3.96	1.97
	$10 \to 8$	4.83	1.17	1.05	5.05
5.	Normal	2.97	3.19	3.26	2.89
	$9 \to 8$	1.29	4.89	4.83	1.31
	$10 \to 8$	4.80	1.35	1.36	4.80
6.[7]	Normal	2.98	3.19	3.10	3.05
	$9 \to 8$	2.98	3.19	3.10	3.05
	$10 \to 8$	2.98	3.19	3.10	3.05
7.	Normal	2.45	3.58	3.43	2.55
	$9 \to 8$	1.24	4.85	4.73	1.29
	$10 \to 8$	4.49	1.52	1.40	4.60
8.	Normal	4.01	2.05	2.34	3.72
	$9 \to 8$	1.97	4.11	4.38	1.63
	$10 \to 8$	4.86	1.25	1.19	4.93

Notes [1]—[6]: See Table II

[7]—Pins 9 and 10 are evidently not bonded to the chip.

Table V—Calculated Input Offsets for Gold-Doped Amplifiers of Table IV.

Input Offset (mV)

UNIT	FF_1 Input Set to $-0.8v$	FF_2 Input Set to $-0.8v$
1	$+ 0.0$	$- 1.0$
2	$+ 0.31$	$- 0.31$
3	$+ 2.05$	$- 1.98$
4	$+ 0.74$	$- 1.25$
5	$- 0.22$	$+ 0.37$
6	$- 0.21$	$+ 0.05$
7	$- 1.13$	$+ 0.88$
8	$+ 1.96$	$- 1.38$

however, so as to short out say 5% of either collector resistor, resulting in an improvement in offset if the 2-mV limit holds for larger samples.

Preliminary Small-Signal Pulse Tests

Dynamic (pulse) testing included both gold-doped and standard-process sense amplifiers. Because of excessive saturation delay, the dynamic tests on the standard samples of Table II were restricted to small-signal pulse tests with the strobe circuitry disabled. Figure 9 shows the typical performance obtained from a sense amplifier made
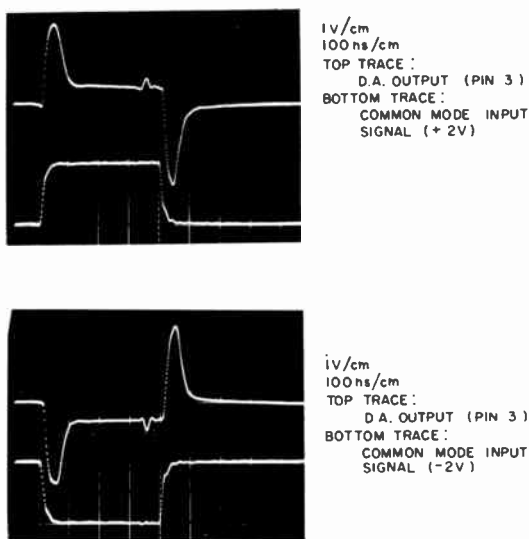
TEST "a" — SHARP RISE TIME INPUT PULSE   TEST "b" — DEGRADED RISE TIME INPUT PULSE

50ns/cm   50ns/cm
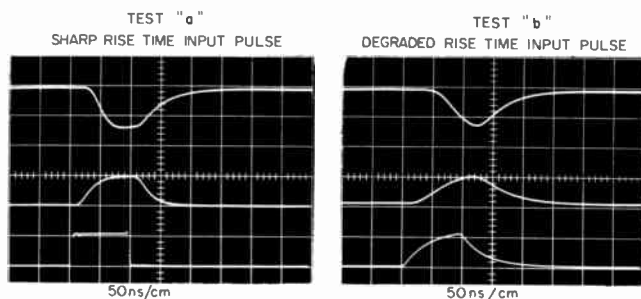
TOP TRACE (BOTH PHOTOS) :- 500mV/cm
  DETECTOR OUTPUT WITH STROBE DISABLED.
MIDDLE TRACE (BOTH PHOTOS) :- 500 m V/cm
  DIFFERENTIAL AMPLIFIER OUTPUT (PIN 3)
BOTTOM TRACE (BOTH PHOTOS) :- 1mV/cm
  SENSE AMPLIFIER INPUT PULSE (PIN 12)

PULSE INPUT (PIN 12)	D A OUTPUT SINGLE ENDED (PIN 3) m V	D A OUTPUT RISE TIME ns	INPUT-D A OUT DELAY ns	DETECTOR		
				AMP (PEAK) m V	RISE TIME ns	OVERALL DELAY ns
1mV SHARP	475	38	29	612	30	48
1mV DEGRADED	440	—	30	530	43	60

Fig. 9—Sense-amplifier pulse test (standard unit No. 3).

with standard processing. Oscillograms showing the response to both sharp- and degraded-rise-time input pulses are included.

Testing of the gold-doped samples included operation of the strobe circuitry. This was made possible due to the gold-doping process whereby the excessive storage time in the saturating transistors, such as the strobe (clamp) transistor, could be limited. Thus, the gold-doped amplifiers could be tested and operated in an entirely realistic manner.

1v/cm
100ns/cm
TOP TRACE :
    D.A. OUTPUT (PIN 3)
BOTTOM TRACE :
    COMMON MODE INPUT
    SIGNAL (+ 2V)

1v/cm
100ns/cm
TOP TRACE :
    D.A. OUTPUT (PIN 3)
BOTTOM TRACE :
    COMMON MODE INPUT
    SIGNAL (−2V)

INPUTS A (PIN 12) AND B (PIN 13) HAVE BEEN
SHORTED TOGETHER FOR THIS TEST.
POSITIVE SUPPLY (NOMINAL + 6 V) HAS BEEN SET TO
+ 8 0V FOR SATURATION PREVENTION.
GOLD DOPED SENSE AMPLIFIER

Fig. 10—Common-mode rejection tests.

Common-Mode Rejection

Preliminary common-mode rejection tests of the gold-doped samples were made. Figure 10 includes oscillograms indicating common-mode rejection. A measure of common-mode rejection is the common-mode rejection ratio, which can be defined as the ratio of differential gain to the common-mode gain. High-frequency common-mode rejection, as evidenced by the sharp peaking at leading and trailing edges of the output pulse, can be readily approximated. High-frequency (~ 10 MHz) common-mode gain (single ended) was measured to be 1.3 and nominal differential mode gain (single ended) is 500, giving a high-frequency common-mode rejection ratio of $500/1.3 = 385$. In a similar

manner, the low-frequency common-mode rejection ratio was found to be $500/0.25 = 2000$.

Pulse Test Apparatus and Arrangement

The test setup of Figure 11 was used in all subsequent pulse testing of the gold-doped sense amplifiers. It provides a means of combining a high-voltage (2 V) common-mode pulse (coupled to both inputs) with

SIGNAL GENERATOR :- RUTHERFORD MODEL B 16
DIGIT GENERATOR :- HEWLETT PACKARD MODEL 214 A
STROBE GENERATOR:- RUTHERFORD MODEL B7 B
OSCILLOSCOPE :- TEKTRONIX TYPE 567 READOUT SAMPLING
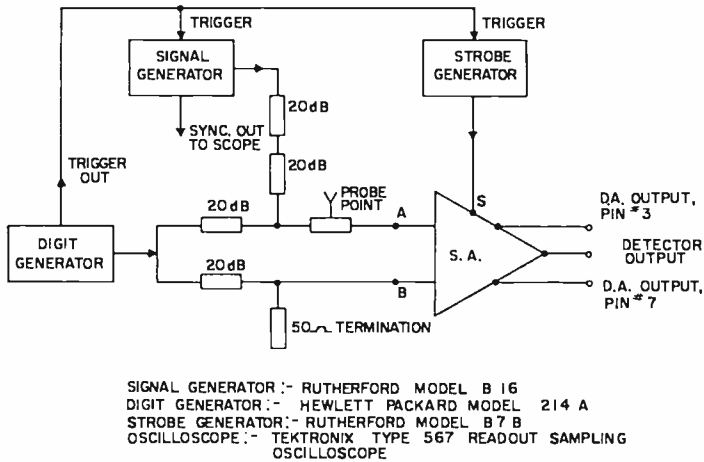   OSCILLOSCOPE

Fig. 11—Pulse test arrangement and apparatus.

a low voltage (1 mV) differential-mode signal pulse (coupled to a single input). The high-voltage common-mode pulse simulates the large digit blast back voltage generated by the laminated-ferrite memory. The low-voltage differential pulse, occurring approximately 0.5 μsec later, simulates the memory signal pulse. A digital readout sampling oscilloscope was used for all measurements and oscillograph recording.

Small-Signal Pulse Tests

Table VI lists the pulse-test results obtained with the gold-doped samples. The digit pulse was not applied for these preliminary tests, in order to avoid interference with the measurements by the digit-pulse tail (decaying exponential), which can be seen in Figure 13. This tail is due to the pulse generator used, and is not a characteristic of the sense amplifier. It is discussed later.

Since the sense amplifier is somewhat sensitive to signal rise time due to the differentiating action of the coupling capacitor, separate

tests were run with input pulses of 3 nsec rise time and 40 nsec rise time. Performance differences were not significant. Input-pulse rise time had to be increased to more than 100 nsec before sense-amplifier sensitivity was significantly reduced.

The start of the strobe pulse was adjusted to coincide with the start of the signal pulse at the sense-amplifier input. Since there is approximately a 35-nsec signal delay (see Table VI) through the two

TABLE VI—Gold-Doped Samples, Pulse Tests

| Unit No. | Input Pulse[1] | | D.A. Single Ended | | | Detector (Strobed) | | |
	Ampli-tude[2] (mV)	Rise Time (ns)	Output[3] (mV)	Rise Time (ns)	Delay (ns)	Ampli-tude[2] (mV)	Rise Time (ns)	Delay[4] (ns)
1.	1	3	560	35	34	910	11	42
	1	40	540	57	38	920	—	39
2.	1	3	500	33	33	820	12	44
	1	40	480	56	37	850	—	41
3.	1	3	450	36	34	850	16	50
	1	40	430	56	38	860	—	43
4.	1	3	510	38	34	920	13	46
	1	40	490	57	39	930	—	41
5.	1	3	540	33	33	950	12	49
	1	40	520	57	37	950	—	47
6.	1	3	480	31	32	920	13	45
	1	40	460	58	35	920	14	46
8.	1	3	500	38	34	900	12	40
	1	40	450	59	35	900	14	39

[1] —100 nsec input pulse width, at "A" input signal only, no digit pulse
[2] —Peak amplitude
[3] —Pin 3
[4] —Overall delay through the sense amplifier

amplifier stages, the strobe (clamp) transistor (Q_{15} in Figure 7) has this period of time to come out of saturation and completely cut off. Without gold doping, the strobe transistor will not completely cut off in 35 nsec. It maintains, due to stored charge, a sufficiently high conductance path to ground so as to short out the amplified signal pulse. For standard-process sense amplifiers (no gold doping), the strobe had to be turned off more than several hundred nanoseconds before the signal pulse, so that the signal could trigger the detector. With the gold-doped amplifiers, however, the 35 nsec proved to be more than adequate.

Figure 12 includes oscillograms showing the small-signal pulse performance of a gold-doped sense amplifier (Unit 1 of Table VI). The middle trace of each photograph, showing the single-ended output of

the D.A., was taken with the strobe disabled, in order to prevent signal distortion caused by the strobe from interfering with the measurements. Strobe distortion is discussed in detail later. However, the
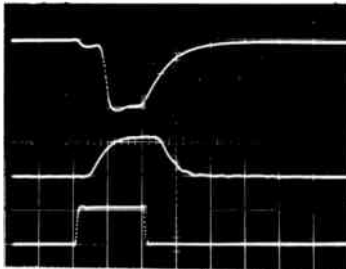
50ns/cm
TOP TRACE : 500 mV/cm
     DETECTOR OUTPUT (STROBED)
MIDDLE TRACE : 500 mV/cm
     D.A. OUTPUT (NO STROBE)
BOTTOM TRACE: 1m V/cm
     1mV INPUT PULSE WITH 3ns RISE TIME
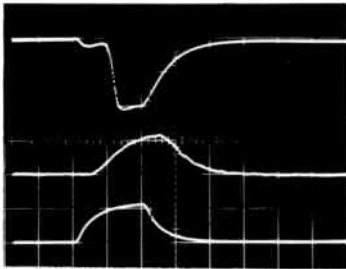
50 ns /cm
TOP TRACE : 500 mV/cm
     DETECTOR OUTPUT (STROBED)
MIDDLE TRACE : 500mV/cm
     D.A. OUTPUT (NO STROBE)
BOTTOM TRACE : 1 mV/cm
     1mV INPUT PULSE WITH 40ns RISE TIME
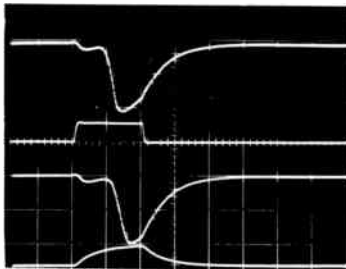
50 ns /cm
TOP PAIR :
   TOP TRACE ' 500mV/cm
       DETECTOR OUTPUT (STROBED)
   BOTTOM TRACE: 1mV/cm
       0.5mV INPUT PULSE WITH 3ns
       RISE TIME
BOTTOM PAIR :
   TOP TRACE ' 500 m V/cm
       DETECTOR OUTPUT (STROBED)
   BOTTOM TRACE ' 1m V/cm
       0.5 mV INPUT PULSE WITH 40ns
       RISE TIME

Fig. 12—Oscillograms of pulse tests of gold-doped sense amplifier (Unit No. 1).

detector output (the top trace) was obtained with the strobe activated.

The top and middle oscillograms of the figure correspond to the tabulated data of Unit 1, but these are typical for all the gold-doped samples tested. The bottom photograph of the figure indicates that substantially full output from the detector is still obtained with 0.5 mV input signal pulses.

Simulated Digit and Sense Input Signals

Oscillograms of the composite test-signal inputs to the sense amplifier, including the digit pulse, are shown in Figure 13. The test set-up was that of Figure 11 described previously. The top photograph in

100 m V/cm
100 ns/cm
DIFFERENTIAL SIGNAL INPUT TO S.A.
  AS SEEN BY SAMPLING SCOPE
TEST CONDITION : 2V DIGIT
  PULSE FOLLOWED BY 10mV
  SIGNAL PULSE

20mV/cm
100 ns/cm
AS ABOVE BUT WITH EXPANDED
  VERTICAL SCALE

50 m V/cm
100 ns/cm
INDIVIDUAL INPUTS TO S A
  ARE SHOWN
BOTTOM TRACE INVERTED IN
  SCOPE FOR CLARITY
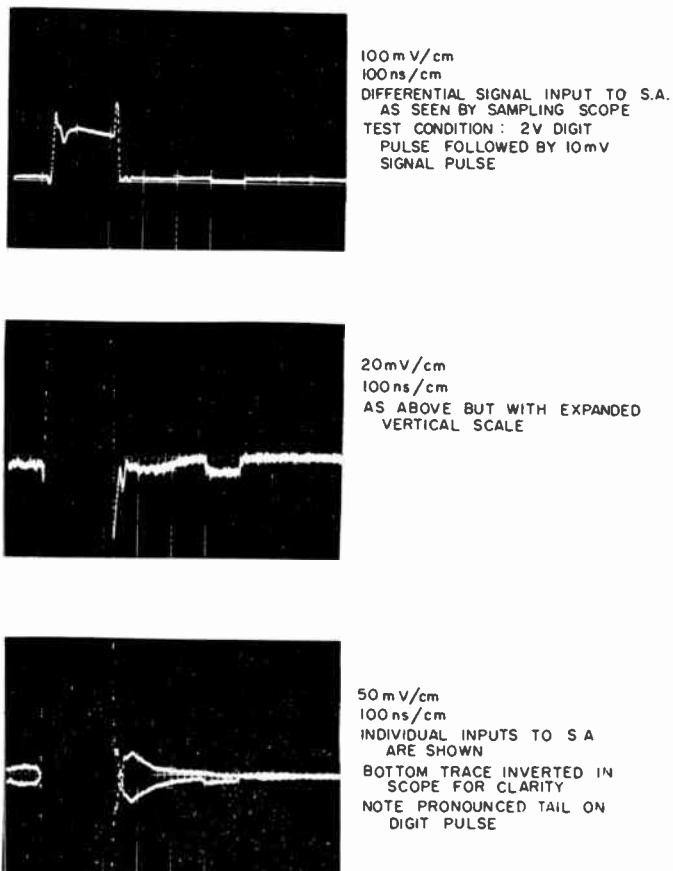NOTE PRONOUNCED TAIL ON
  DIGIT PULSE

Fig. 13—Oscillograms of differential input signal tests.

Figure 13 shows the differential digit-signal input to the sense amplifier as seen by the sampling oscilloscope. If the common-mode coupling of the digit pulse to the sense amplifier inputs were perfect, then no differential signal would be observable. However, due to the slightly unequal attenuation and path delays occurring in the two separate paths from the digit pulse generator to the two sense amplifier inputs, somewhat unequal digit pulses occur at the inputs.

The middle photograph of Figure 13 shows, on an expanded verticle scale, the differential digit signal input. A negative 10-mV simulated sense-signal pulse can also be seen. Note the tail immediately following the large digit pulse. The bottom photograph, showing the individual sense amplifier inputs, shows this tail more clearly. It is obviously a characteristic of the digit pulse generator used and would not be present when the amplifier is used in conjunction with the laminated-ferrite memory.
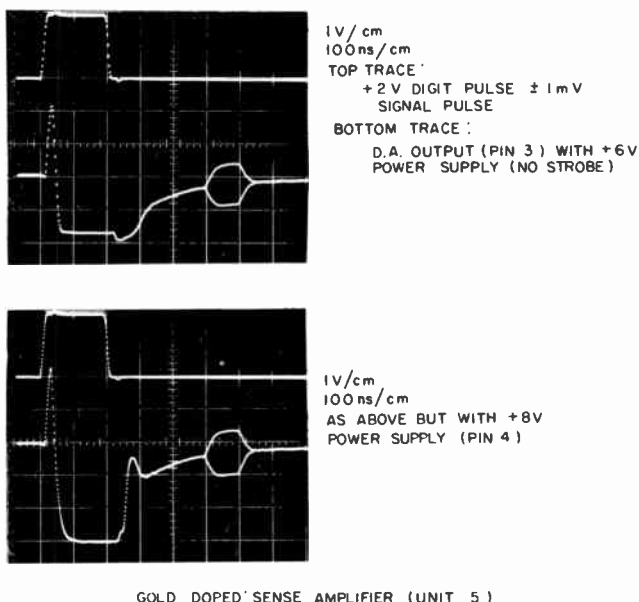

I V / cm
100 ns / cm
TOP TRACE
   + 2 V DIGIT PULSE ± I m V
     SIGNAL PULSE
BOTTOM TRACE :
   D.A. OUTPUT (PIN 3) WITH + 6 V
   POWER SUPPLY (NO STROBE)

I V/cm
100 ns/cm
AS ABOVE BUT WITH + 8V
POWER SUPPLY (PIN 4)

GOLD DOPED SENSE AMPLIFIER (UNIT 5)

Fig. 14—Oscillograms of differential amplifier saturation tests.

Saturation Effects

Pulse tests were made on the gold-doped sense amplifiers to determine the effects of second-stage saturation. The collector resistors of this stage are returned to a separate positive power supply at pin 4. Nominal supply voltage is 6 volts. With this supply voltage, the second stage will not saturate at the small sense-signal levels expected from the laminated-ferrite memory. It will, however, saturate at the high digit-pulse levels, as is readily seen in the oscillograms of Figure 14. The top photograph shows the effects of second-stage saturation with the supply set to 6 volts. The lower photograph illustrates how saturation can be avoided by setting the supply to 8 volts. This costs an additional 50 mW in amplifier power dissipation with the present de-

sign. Whether saturating or non-saturating operation is employed will depend upon the memory-cycle time requirements. If maximum sense amplifier performance is required, non-saturating operation should be used.

Strobe Distortion

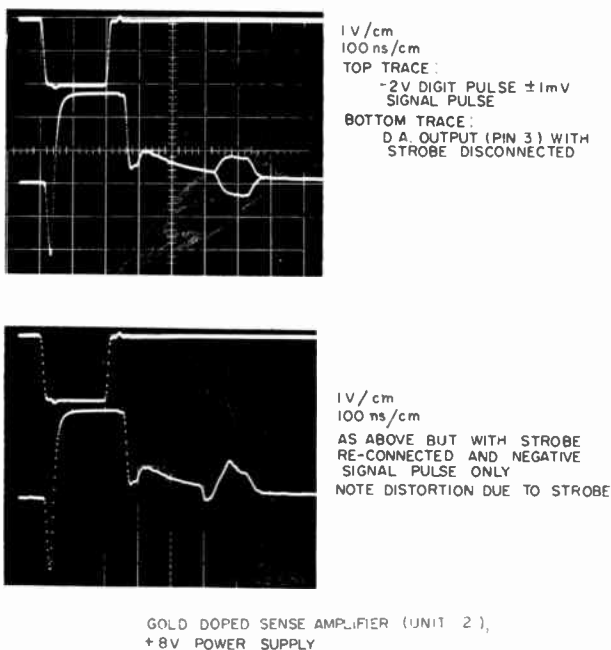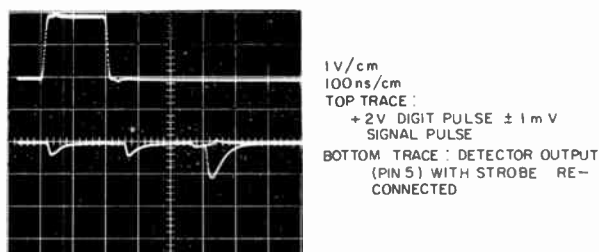Figure 15 shows oscillograms indicating the effect of strobe-circuit

1 V/cm
100 ns/cm
TOP TRACE :
    ⁻2 V DIGIT PULSE ± 1 mV
    SIGNAL PULSE
BOTTOM TRACE :
    D. A. OUTPUT (PIN 3) WITH
    STROBE DISCONNECTED

1 V/cm
100 ns/cm
AS ABOVE BUT WITH STROBE
RE-CONNECTED AND NEGATIVE
SIGNAL PULSE ONLY
NOTE DISTORTION DUE TO STROBE

GOLD DOPED SENSE AMPLIFIER (UNIT 2),
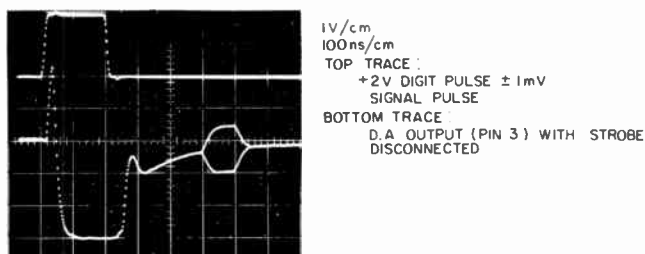+ 8 V POWER SUPPLY

Fig. 15—Oscillograms of strobe circuit tests.

operation on the pulse waveforms. The complete pulse test pattern including the 2-volt digit pulse followed by the 1-mV signal pulse is used. The output is taken at pin 3 (D.A. output). The strobe has been disconnected for the top photograph. The bottom photograph shows the waveform distortion with the strobe reconnected. This distortion is due to either the collector–base capacitance (~ 4 pF) or stored charge in the strobe transistor. Increasing the size of the coupling capacitor reduces this effect. The 22-pF capacitance used represents about the minimum usable without introducing too severe distortion (enough to trigger the detector without a signal pulse) and without making amplifier sensitivity too rise-time dependent. Increasing the capaci-

tance would result in less marginal amplifier performance at the cost of increased chip area.

Detector Discrimination

The detector output, for the same input pulse test pattern as above, is shown in Figure 16. The top photograph again shows the output

IV/cm
IOOns/cm
TOP TRACE :
    +2V DIGIT PULSE ± ImV
    SIGNAL PULSE
BOTTOM TRACE :
    D.A OUTPUT (PIN 3) WITH STROBE
    DISCONNECTED

IV/cm
IOOns/cm
TOP TRACE :
    + 2V DIGIT PULSE ± I m V
    SIGNAL PULSE
BOTTOM TRACE : DETECTOR OUTPUT
    (PIN 5) WITH STROBE   RE—
    CONNECTED

GOLD DOPED SENSE AMPLIFIER (UNIT I),
+ 8V POWER SUPPLY

Fig. 16—Oscillograms of combined digit pulse and signal pulse tests.

from the D.A. with the strobe disconnected. The bottom oscillograph shows the detector output (pin 5) with the strobe reconnected. Discrimination between the 1-mV "one" and "zero" simulated sense signals is evident. However, a partial response caused by the digit pulse is also present. The high (2-volt) positive-going transitions of the simulated digit pulse (at the input) are not being entirely absorbed in the strobe circuit.

The above test condition represents a very severe test. Such high dv/dt's as with the simulated digit pulse would probably not be encountered in use. If they are to be allowed for, however, several remedies are possible. (1) The base drive on the strobe transistor can be

increased, driving it harder into saturation. This would probably increase the distortion problem discussed previously and necessitate, a somewhat larger coupling capacitance. (2) Alternatively, or in addition to the preceding, the detector threshold could be increased. This would decrease the sense amplifier sensitivity to noise at the cost of lower overall gain. The gain of the D.A. stages could be increased to offset this.

The possible circuit modifications suggested here as a means of reducing the amplifier's sensitivity to very high dv/dt digit pulses are primarily a matter of readjusting resistor ratios. The basic circuit configuration, which has been demonstrated to be sound, does not require alteration.

CONCLUSIONS

A number of techniques for the design of an integrated, high-gain (60 dB), high-sensitivity (1 mV) sense amplifier have been explored. Several techniques including the somewhat novel approach of using feedback by thermal means and a more conventional approach using a-c coupling have been proven feasible.

A sense amplifier utilizing thermal feedback for automatic input offset correction was constructed using discrete components. Thermal feedback takes advantage of the temperature dependence of the base–emitter characteristic of the amplifier input differential transistors to achieve automatic balancing. The effective input imbalance of a high-gain (60 dB) breadboarded amplifier was controlled to within ± 30 microvolts. This represents an improvement by a factor of 100 over the identical amplifier with the thermal-feedback loop open. In a high-gain dc-coupled amplifier, the effective input imbalance of the amplifier (typically 5 mV) represents one of the major design problems.

A number of tests with available integrated structures indicated that offset corrections of 2 mV could be readily obtained with less than 50 mW dissipation. With optimum geometries, the dissipation required should be considerably less. Further study of the dissipation requirements for establishing thermal gradients in silicon integrated structures is required to verify the practicality of incorporating thermal feedback into a monolithic amplifier.

A high-gain (60 dB for 1 mV sensitivity) ac-coupled sense amplifier was designed, tested with discrete parasitic components, and subsequently integrated. The amplifier has a bandwidth of 10 MHz, which is somewhat better than the design specification (3 MHz). The

amplifier measures 70 × 75 mils and dissipates 135 milliwatts. It is packaged in a 14-lead flat pack.

The sense amplifier includes a two-stage differential amplifier capacitively coupled to a strobed detector. It combines 1-mV signal sensitivity with a ± 2-volt digit pulse capability. Amplifier recovery time from digit overdrive is less than 100 nsec. Because of internal inversion logic circuitry, the sense amplifier can be used with a split-sense-line organized laminated-ferrite memory.

The sense amplifier has been fabricated with standard, n+ pocket, and gold-doped processes. Gold doping is required for limiting storage time in the saturating strobe and detector transistors. The gold-processed amplifiers were extensively pulse tested and performed substantially as initially specified. Input offsets for the samples tested were 2 mV or less. Typical amplifier performance is tabulated as follows.

```
Gain  (overall)  ............Approximately 60 dB  (0.9-V output with
                              0.5 — 1.0 mV input)
Bandwidth  ................10 MHz (differential stages)
Digit Pulse, Maximum  .....± 2 V
Threshold  ................0.5 — 1.0 mV
Output Voltage  ............0.9 V
Temperature Range  ........10°C to 65°C
```

ACKNOWLEDGMENTS

AN EXPERIMENTAL PULSED CdS LASER CATHODE-RAY TUBE

BY

F. H. NICOLL

RCA Laboratories
Princeton, N. J.

Summary—A laser cathode-ray tube is described that has aluminized CdS crystals pumped by the pulsed electron beam. The tube configuration is of conventional design; the laser beam is fan shaped with a spread of 180° in one direction and less than 10° in the other. Although there are several crystals on the faceplate, the present tube is suitable only for single-spot low-duty-cycle operation. However, in theory it should be possible, by the use of many carefully aligned crystals on the tube face, to make a scanable directional cathode ray tube having a wide horizontal viewing angle and a narrow vertical angle.

INTRODUCTION

ELECTRON-beam-pumped lasers have now been reported with wavelengths from 3100 Å in the UV to 8 microns in the IR.[1] These results have been obtained in demountable vacuum systems at a temperature of 77°K or lower and in some cases, e.g., CdS, with high power efficiency. Lasing of CdS on a room-temperature substrate was first reported by the author[2] for a total-internal-reflection mode.[3] Published results of other workers using Fabry–Perot modes, which in general have higher thresholds, indicate that lasing in CdS has been obtained only up to 250°K, even when the substrate was maintained at a very much lower temperature.

A laser cathode-ray tube was proposed,[4] but not built, using the Fabry–Perot mode of operation from a staircase assembly of semiconductor slices, with viewing taking place at right angles to the tube axis. This arrangement is very cumbersome and suffers from the many scanning, viewing, and construction problems associated with even simple phosphor screens viewed on the bombarded side.

[1] M. I. Nathan, "Semiconductor Lasers," *Appl. Opt.*, Vol. 5, No. 10, p. 1514, Oct. 1966.

[2] F. H. Nicoll, "Room-Temperature Lasing of CdS Under Pulsed Electron Bombardment," *Appl. Phys. Letters*, Vol. 10, No. 3, p. 69, Feb. 1967.

[3] F. H. Nicoll, "Far-Field Pattern of Electron Bombarded Semiconductor Lasers," *Proc. IEEE (Letters)*, Vol. 55, No. 1, p. 114, Jan. 1967.

[4] B. Lax, "Scanatron—A Scanning Beam Semiconductor Laser," *Solid State Design*, Vol. 6, No. 3, p. 19, Mar. 1965.

The total-internal-reflection mode of lasing produces a fan-shaped beam spreading 360° in the horizontal plane and less than 10° in the vertical direction. Such an emission pattern lends itself to a conventional cathode-ray-tube structure viewed, as usual from the front of the faceplate. This, in conjunction with low-voltage room-temperature operation, makes a sealed-off laser cathode-ray tube possible.

ELECTRON–OPTICAL DESIGN

Laser action under electron-beam pumping requires a higher current density than is usually obtained in commercial cathode-ray tubes. Even the spot size in a projection kinescope such as the 5-inch-diameter 5TP4 tube does not give adequate current density at the 1 mA of current available. Smaller focused spots are most readily obtained by moving the faceplate very near the final focusing lens of the tube. This is the case whether electrostatic or electromagnetic focusing is used. The proximity of the samples to the final lens makes a very short tube possible, but it also reduces the screeen area over which adequate focus can be maintained.

The low-threshold and low-voltage operation of the internal-reflection-mode laser makes an electrostatic-focus tube possible even at operating voltages below 25 kV. The present tube was designed for single-spot pulsed operation using electrostatic focus, but the addition of a final magnetic focusing lens gives considerably improved results.

LASER-CRYSTAL PROPERTIES AND SCREEN PREPARATION

The 360° internal-reflection lasing mode has the lowest threshold because total reflection is virtually lossless and the absorption losses in the crystal are small.[5] In order to keep these losses as small as possible, the electron-beam-excited depth should be comparable to the crystal thickness. To do this at beam voltages under 25 kV, the thinnest possible crystals are used; in practice, they are about 2 microns thick and are grown as platelets parallel to the *c*-axis. Such crystals have the further advantage at room temperature of providing good heat conduction to the substrate. These thin crystals are cleaved about 20-100 microns wide parallel to the *c*-axis and perdendicular to the grown face, to give the necessary rectangular-cross-section cavity. The length of the strips is unimportant, but very thin crystals as grown from the vapor are seldom longer (parallel to the *c*-axis) than 1.0 mm.

---

⁵ F. H. Nicoll, "Far-Field Patterns of Electron-Beam Pumped Semiconductor Lasers," *Jour. Quantum Electronics*, Vol. QE-4, No. 4, p. 198, April 1968.

For room-temperature operation the crystals must be in intimate contact with the faceplate for adequate cooling. Aluminization is desirable to eliminate cathode light and to increase the light emitted in the forward direction. It is also useful in assuring that charging of the samples does not occur. It became apparent that conventional aluminizing on a plastic film covering the samples was not possible because of the high-temperature air bake required to burn out the plastic. This air bake did not visibly damage the crystals, but raised the room-temperature threshold for lasing.

The technique adopted for making the screen of single-crystal platelets is as follows. A one-inch-diameter faceplate of sapphire or glass is coated with transparent tin oxide except for a $\frac{1}{8}$-inch-diameter area in the center. This area is charged by a corona discharge, and the cleaved samples, placed in position, are held by the charge. The plate is then immersed in water with the samples still in place. 1000-Å-thick dots of aluminum $\frac{1}{4}$ inch in diameter are vacuum evaporated through a mask onto a microscope slide previously coated with sodium chloride by evaporation. Each aluminum dot is floated onto the surface of the water over the crystal samples by dissolving the sodium chloride during submersion of the glass slide. On removing the water the aluminum dot settles down over the crystals and on being dried, firmly holds the crystals against the faceplate. As in conventional phosphor aluminizing, the aluminum drapes around the sides of the crystal making an excellent reflector for the laser light. It is important that the aluminum film should not make optical contact with the CdS crystals since it has been shown[3] that the lasing threshold is raised considerably, and lasing may actually cease, due to the reflection loss at the Al-CdS interface. This loss, though small, is much greater than that for total internal reflection. The loss in the aluminum is prevented in the present case by the oxidation that takes place on the under side of the aluminum film while it floats on the water. Al_2O_3 (refraction index $n = 1.75$) serves to isolate the aluminum optically and allow total reflection to occur at the $CdS-Al_2O_3$ interface. However, the refraction index of Al_2O_3 is excessively high, and a better result can be obtained by using a 300-Å layer of MgF_2 ($n = 1.34$) underneath the floated-on 1000-Å-thick aluminum film.

Figure 1 is a photograph of some of the aluminized crystals taken through the faceplate. Surrounding the crystals is a dark area where the aluminum slopes down from the back of the crystal to the faceplate at an angle of about one or two degrees to the faceplate. The intimate contact between the crystals and the faceplate is evidenced by the interference pattern visible under some of the crystals. The crystals
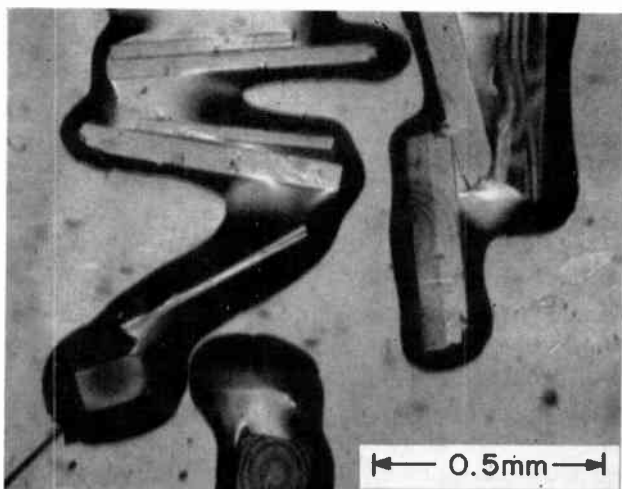
Fig. 1—Aluminized crystals photographed through the faceplate of the cathode-ray tube.

and faceplate are air baked at 150°C and, after sealing to the tube by any of several well-known techniques, can be vacuum baked up to about 150°C without affecting laser threshold.

DESCRIPTION AND OPERATION OF TUBES

Figure 2 is a photograph of a completed tube, showing the 5TP4-type gun, and the aluminized crystals in the center of the faceplate.
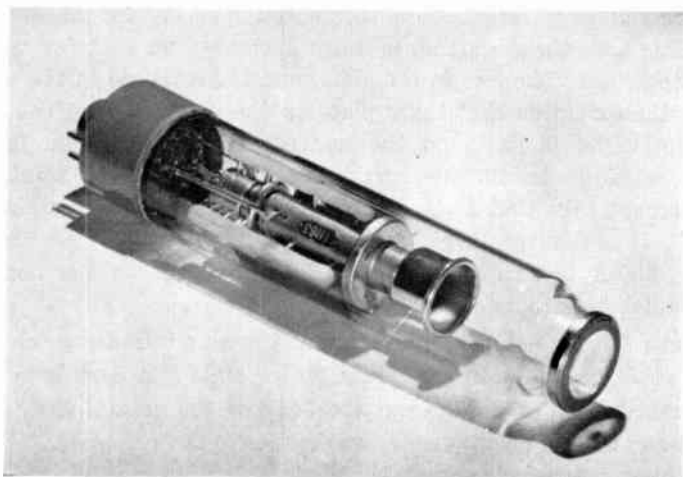

Fig. 2—Photograph of laser cathode-ray tube showing electron gun and aluminized samples in the center of the faceplate.

Between the end of the gun and the faceplate a transparent tin oxide coating serves as the final anode and cylindrical focusing lens. For a vertical crystal, the fan-shaped laser beam is horizontal and is emitted from the crystal through the faceplate.

The usual operating voltages are applied to the 5TP4 gun except that the voltage ratio between the last two anodes is about 10. The electron beam, which can be positioned by a small magnet or deflection coil, is pulsed with 50-volt 50-nsec pulses on the grid at a rate of 10-100 Hz. This low duty cycle is used to reduce heating and minimize a slow degradation in laser output. At 10 Hz, a bombarded spot on one crystal

Fig. 3—Photomicrograph of one CdS crystal emitting laser light at corners.

will degrade gradually until, after one or two hours, lasing ceases. The reason for this degradation is not known at the present time.

Figure 3 is a photomicrograph showing the laser light being emitted from two intense spots at the cleaved edges of the vertical CdS crystal. Light from the irregular, diffuse, electron-bombarded area is visible in the middle of the crystal. The horizontal fan-shaped laser beam is emitted in the plane containing these two spots and the perpendicular to the plane of the CdS crystal and faceplate. The emitted laser light is strongly polarized with the E vector horizontal and perpendicular to the c-axis of the crystal.

Figure 4 shows a portion of the interference pattern produced by the two coherent beams from a vertical cleaved CdS sample 21 microns wide. The photographic film was located in front of the tube at a distance of 8 cm from the crystal and approximately parallel to its

Fig. 4—Photograph interference pattern of laser beams (taken at about 8 cm from crystal).

broad face. The light from the lasing crystal was allowed to fall directly on the film without interposition of a lens. The interference lines are present only when the sample is lasing, and their calculated and measured spacing is in good agreement for light of the laser wavelength. The narrow angle of emission in the vertical plane is also evident from the photograph, since the vertical length of the interference lines corresponds to an angle of about 8°. These interference lines are only observed sharply near threshold; at higher excitation levels, they merge to give a uniform 180° emission pattern with the same narrow vertical angle.

Figure 5 shows the spectrum of the light emitted from the CdS crystals below laser threshold. The narrow lasing line is shown in the spectrum of Figure 6, which was taken just above threshold. This figure also shows the incoherent broad peak. The current density for lasing is about 4 amperes per cm^2, and it is possible to obtain lasing down to an accelerating voltage of 10 kV if a conventional magnetic focus coil is used to supplement the electrostatic focus. The rise and decay time of the laser light is less than 15 nsec, which is the limit of
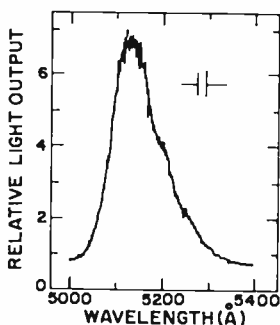

Fig. 5—Broad emission spectrum of CdS crystal operated below laser threshold near room temperature.

measurement on the oscilloscope. Power efficiency is about 0.5%, giving a peak pulsed output of about 120 mW of green laser light.

CONCLUSION

A pulsed-laser cathode-ray tube has been described that uses CdS single crystals operating at room temperature. The present tube, although containing a number of crystals is really only suitable for
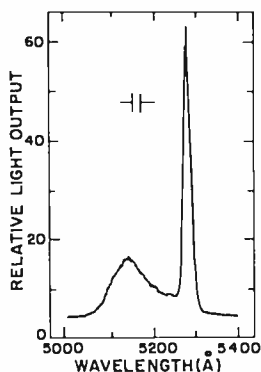

Fig. 6—Emission spectrum of same CdS crystal as in Figure 5 operated above laser threshold showing narrow laser line.

single-spot, low-duty-cycle operation. A scanning type tube would require many more aligned crystals operated with synchronized pulsing. While technically possible, the active area of such a screen scanned at a reasonable angle would only be about $\frac{1}{4}$ inch. A tube with horizontal rows of cleaved crystals with their c-axes vertical would all emit a polarized 180° horizontal fan shaped beam of about 10° vertical angle. Thus, for the first time a cathode-ray tube having a directional emission pattern is a technical possibility.

ACKNOWLEDGMENTS

ELECTROMAGNETIC WAVE PROPAGATION
IN SUPERCONDUCTORS

By

P. Bura

RCA Defense Communications Systems Division
Princeton, N. J.

Summary—Propagation of electromagnetic waves in superconductors differs from propagation in normal metals by the presence of the Meissner effect, i.e., the expulsion of magnetic flux from the superconducting bulk. Ginzburg and Landau have proposed a theory to account for this effect. This paper shows that, to the extent that charge effects can be neglected, the equilibrium Ginzburg–Landau equations can be used together with Maxwell's equations to derive the solution to EM wave propagation in superconductors.

The transmission line structures that have negligible fields along the boundary with open space and have transverse dimensions of the order of penetration depth exhibit appreciable variation of the propagation constant with magnetic field (or current). This makes them suitable for distributed traveling-wave parametric amplifiers.

INTRODUCTION

PROPAGATION of electromagnetic waves in superconductors differs from propagation in normal metals by the presence of the Meissner effect, i.e., the expulsion of magnetic flux from the superconducting bulk. This is somewhat similar to the "skin-depth" phenomenon of normal metals at microwave frequencies, except that the magnetic-field penetration depth in superconductors is several orders of magnitude smaller than the skin depth and is frequency independent, i.e., the field expulsion takes place in both steady and time-varying fields. F. and H. London[1] have proposed the following equation to account for the Meissner effect:

$$\nabla \times \mathbf{J} = -\frac{C}{4\pi\lambda_L} \cdot \mathbf{H}, \tag{1}$$

where λ_L is the magnetic-field penetration depth, J is the current density, and H is the magnetic field.

---

[1] F. and H. London, "The Electromagnetic Equations of the Supraconductor," *Proc. Royal Soc.*, Vol. A149, p. 71, Mar. 1935.

Although the London theory successfully accounts for many super-conducting phenomena, it suffers from the "local" character of its current-vector potential $(J\text{-}A)$ relation. In addition, it does not account correctly for the surface energy and for the field or current destruction of superconductivity.

To overcome these difficulties, Ginzburg and Landau (G-L) proposed a phenomenological theory[2] based on the Landau–Lifshits theory of phase transitions. Ginzburg and Landau assume an order parameter ψ— a macroscopic wave function for superconducting pairs— that can be normalized to their density

$$|\psi|^2 = n_s .$$

In accordance with the theory of phase transitions, they express the free-energy density of a superconductor as an even power series of the order parameter:

$$F_{s0} = F_n - \alpha|\psi|^2 + \frac{1}{2}\beta|\psi|^4 , \tag{2}$$

where F_{s0} = zero-field free-energy density, F_n = free-energy density of the normal metal, and α, β = material constants. In the presence of a magnetic field, the expression becomes

$$F_{sH} = F_{s0} + \frac{H^2}{8\pi} + \frac{1}{2m}\left| -ih\nabla\psi - \frac{e}{c}A\psi \right|^2 , \tag{3}$$

where the last term is analogous to the kinetic energy term in the electron-field interaction of quantum mechanics. The inclusion of the vector potential makes the expression gauge invariant.

The equilibrium condition is then obtained by varying F_{sH} with respect to ψ^* and A. This leads to two coupled equations;

$$\frac{1}{2m}\left(-ih\nabla - \frac{e}{c}A \right)^2\psi - \alpha\psi + \beta|\psi^2|\psi = 0 , \tag{4}$$

$$J = \frac{ieh}{2m}(\psi\nabla\psi^* - \psi^*\nabla\psi) - \frac{e^2}{mc}|\psi|^2A . \tag{5}$$

[2] V. L. Ginzburg and L. D. Landau, "On the Theory of Superconductivity," *J.E.T.P. USSR*, Vol. 20, p. 1064, 1950.

The validity and usefulness of these equations have been verified experimentally. However, they are equilibrium expressions and, therefore, cannot be applied a priori to time-varying problems such as EM wave propagation.

TIME-DEPENDENT GINZBURG–LANDAU THEORY

Gorkov has shown[3] the validity of Ginzburg–Landau equations near the critical temperature, T_c, from microscopic considerations, by identifying the order parameter with the energy-gap function. He pointed out that the order parameter must have a time-dependent phase given by

$$\psi = \psi_m \exp\left\{ -\frac{2i\mu t}{\hbar} \right\}, \tag{6}$$

where μ is the chemical potential. Josephson[4] first noted that this time dependence leads to the a-c tunneling current.

Starting with this basic time dependence of the order parameter, Anderson, et al,[5] have derived, phenomenologically, an additional time-dependent equation to supplement the static G-L equations. Stephen and Suhl[6] obtained a similar expression, valid in the neighborhood of T_C, from the microscopic Barden–Cooper–Schriefer (BCS) theory. Suhl has also shown[7] that the additional time-dependent equation can be derived by forming a Lagrangian from the G-L free-energy expression by addition of

$$\Delta L = -\left[\frac{1}{2mV^2} \left| \left(\hbar\frac{\partial}{\partial t} + 2ic\varphi \right) \psi \right|^2 + \frac{E^2}{8\pi} \right], \tag{7}$$

[3] L. P. Gorkov, "On the Energy Spectrum of Superconductors," *Sov. Phys. JETP* (Transl.), Vol. 34(7), p. 505, 1958.

[4] B. D. Josephson, "Possible New Effects in Superconductive Tunneling," *Phys. Letters*, Vol. 1, p. 251, 1 July 1962; "Coupled Superconductors," *Rev. Mod. Phys.*, Vol. 36, p. 216, Jan. 1964.

[5] P. W. Anderson and N. R. Werthamer, "An Additional Equation in the Phenomenology of Superconductivity: Resistive Effects," *Phys. Rev.*, Vol. 138A, p. A-1157, 17 May 1965.

[6] M. J. Stephen and H. Suhl, "Weak Time Dependence in Pure Superconductors," *Phys. Rev. Lett.*, Vol. 13, p. 797, Dec. 1964.

[7] H. Suhl, "Inertial Mass of a Moving Fluxoid," *Phys. Rev. Lett.*, Vol. 14, p. 226, Feb. 1965.

where

φ = electric potential,
$\mathbf{E}$ = electric field intensity,
$V = V_F/\sqrt{3}$ and V_F is the Fermi velocity.

The full Lagrangian can then be written

$$L = -\alpha|\psi|^2 + \frac{1}{2}\beta|\psi|^4 + \frac{1}{2m}\left| -i\hbar\nabla\psi - \frac{e}{c}\mathbf{A} \right|^2$$
$$- \frac{1}{ZmV^2}\left| -i\hbar\frac{\partial\psi}{\partial t} + e\varphi\psi \right|^2 + \frac{1}{8\pi}(H^2 - \mathbf{E}^2) \tag{8}$$

We can now follow the standard variational procedure to obtain the minimizing relations using the Lagrangian differential equation

$$\frac{\partial L}{\partial q} - \frac{\partial}{\partial x}\left[\frac{\partial L}{\partial\left(\dfrac{\partial q}{\partial x}\right)}\right] - \frac{\partial}{\partial t}\left[\frac{\partial L}{\partial\left(\dfrac{\partial q}{\partial t}\right)}\right] = 0 ,$$

where q is a dynamic variable. Taking $\mathbf{A}$, ϕ, ψ and ψ^* as the independent variables, we arrive at the relations

$$\frac{1}{2m}\left(-i\hbar\nabla - \frac{e}{c}A \right)^2\psi - \frac{1}{2mV^2}\left(-i\hbar\frac{\partial}{\partial t} + e\varphi \right)^2\psi - \alpha\psi + \beta|\psi|^2\psi = 0 , \tag{9}$$

$$\nabla^2\phi + \frac{1}{c}\frac{\partial}{\partial t}(\nabla\cdot A) = -4\pi\rho , \tag{10}$$

$$\nabla^2\mathbf{A} - \nabla\cdot(\nabla\mathbf{A}) - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} - \frac{1}{c}\frac{\partial}{\partial t}(\nabla\varphi) = -\frac{4\pi}{c}\mathbf{J} , \tag{11}$$

where J and ρ are superconducting current and charge densities defined by

$$J = \frac{ie\hbar}{2m}(\psi\nabla\psi^* - \psi^*\nabla\psi) - \frac{e^2}{mc}A|\psi|^2 , \tag{12}$$

$$\rho = \frac{1}{V^2} \left[\frac{ieh}{2m} \left(\psi^* \frac{\partial \psi}{\partial t} - \psi \frac{\partial \psi^*}{\partial t} \right) - \frac{e^2}{m} \varphi |\psi|^2 \right]. \tag{13}$$

We can see that the first static G-L equation [Equation (4)] has been modified by the additional term

$$- \frac{1}{2m V^2} \left(-i\hbar \frac{\partial}{\partial t} + e\varphi \right)^2 \psi .$$

The current expression [Equation (5)] remains unaltered. However, an expression for the charge must now be taken into account. Equations (10) and (11) are the generalized Maxwell's equations. The equations are gauge invariant and the continuity equation

$$\nabla \cdot J + \frac{\partial \rho}{\partial t} = 0$$

follows from the imaginary part of Equation (9).

WAVE PROPAGATION

We now consider the case where only the phase of the order parameter is a function of time. The theory, however, is also applicable to cases where the modulus of the order parameter, ψ_m, becomes a function of time, as in the vortex motion in the mixed state of type-II superconductors.

As already pointed out, the phase of the order parameter is related to the chemical potential of the superconductor. In the presence of electrostatic potential and kinetic energy of Cooper pairs it becomes

$$\mu = \mu_c + e\varphi + \frac{1}{2m} \left| \left(\nabla - \frac{ie}{c} A \right) \psi \right|^2, \tag{14}$$

where μ_c is the material constant for a given temperature. The last term introduces further nonlinearity into the Ginzburg–Landau equations.

We now define a general scalar phase function, W, of the order parameter

$$\psi = \psi_m \exp \left\{ \frac{ie}{\hbar c} W \right\} .$$

The time-dependent Ginzburg–Landau equations can now be written

$$- \frac{\hbar^2}{2m} \nabla^2 \psi_m + \frac{e^2}{2mc^2} \psi_m |A - \nabla W|^2$$
$$+ \frac{e^2}{2mV^2} \left| \varphi + \frac{1}{c} \frac{\partial W}{\partial t} \right|^2 - \alpha \psi_m + \beta \psi_m^3 = 0 ,$$

$$(15)$$

$$\mathbf{J} = \frac{e^2}{mc} \psi_m^2 (A - \nabla W) .$$

$$(16)$$

$$\rho = - \frac{e}{mV^2} \psi_m^2 \left(\varphi + \frac{1}{c} \frac{\partial W}{\partial t} \right) .$$

$$(17)$$

The continuity equation, which follows from Equation (9), now becomes

$$\nabla^2 W - \frac{1}{V^2} \frac{\partial^2 W}{\partial t^2} + \frac{2}{\psi_m} \nabla W \cdot \nabla \psi_m = \frac{c}{V^2} \frac{\partial \varphi}{\partial t} + \nabla \cdot A + \frac{2}{\psi_m} A \cdot \nabla \psi_m .$$

$$(18)$$

It has become a wave equation governing the propagation of the phase function of the order parameter.

Equations (16) and (17) suggest a new set of potentials that would simplify the time-dependent Ginzburg–Landau equations. Putting

$$A' = A - \nabla W \quad \text{with} \quad \nabla \cdot A' = 0 .$$

and

$$\phi = \varphi + \frac{1}{c} \frac{\partial W}{\partial t} .$$

in Equations (16) and (17) and substituting in Equations (10) and (11), respectively, we get

$$\nabla^2 \phi - \frac{c^2}{V^2} \frac{\psi^2}{\lambda_o^2} \phi = 0 ,$$

$$(19)$$

and
$$\nabla^2 \mathbf{A}' - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}'}{\partial t^2} - \frac{\psi}{\lambda_o^2} \mathbf{A}' = 0 , \qquad (20)$$

where λ_o is the zero-field penetration depth. Here Ginzburg–Landau normalization of the order parameter was used. Using the same normalization, Equation (15) can be written

$$- \frac{\lambda_o^2}{\kappa^2} \nabla^2 \psi + \left(\frac{A'}{\sqrt{2} H_c \lambda_o} \right)^2 \psi + \frac{c^2}{V^2} \left(\frac{\phi}{\sqrt{2} H_c \lambda_o} \right)^2 \psi - \psi + \psi^3 = 0 ,$$
$$(21)$$

where κ is the coupling constant and H_c the bulk critical field. The continuity equation now leads to the relation between the potentials;

$$\frac{\partial \phi}{\partial t} = - \frac{2V^2}{c\psi} A' \cdot \nabla \psi. \qquad (22)$$

The fields are invariant to the above potential transformation;

$$\mathbf{H} = \nabla \times \mathbf{A} = \nabla \times \mathbf{A}' ,$$

and

$$\mathbf{E} = - \left(\nabla \phi + \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} \right) = - \left(\nabla \phi + \frac{1}{c} \frac{\partial \mathbf{A}'}{\partial t} \right).$$

Equations (19)-(21) represent a time-dependent set of Ginzburg–Landau equations.

As with the normal metals, because of the smallness of the Debye screening length, the charge is very small. This allows further simplification of the equations by assuming the charge to be zero. Equation (19) then becomes Poisson's equation, while Equations (20) and (21) reduce to the equilibrium Ginzburg–Landau form:

$$\nabla^2 \phi = \nabla^2 \psi = 0 ,$$

$$\nabla^2 \mathbf{A}' - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}'}{\partial t^2} - \frac{\psi^2}{\lambda_o^2} \mathbf{A}' = 0 , \qquad (23)$$

$$- \frac{\lambda_o^2}{\kappa^2} \nabla^2 \psi + a^2 \psi - \psi + \psi^3 = 0 ,$$

where
$$a = \frac{A'}{\sqrt{2}H_r\lambda_o}.$$

One can say, then, that to the extent that the charge effects are negligible, the equilibrium equations can be applied to the time-varying problems.

One-Dimensional Case

We assume an EM wave propagating in the Z direction with a propagation constant γ, with no variation in the Y direction. With periodic variation in time and along the Z direction, this becomes a one-dimensional problem, with quantities varying only in the X direction. Thus,

$$\frac{\partial}{\partial t} \equiv i\omega, \quad \frac{\partial}{\partial z} \equiv -i\gamma, \quad \text{and } \frac{\partial}{\partial y} \equiv 0.$$

From the continuity equation (Equation (22)), we see that A' must have both X and Z components ($\nabla \psi$ is an X vector):

$$i\omega\phi = -\frac{2V^2}{c\psi}A'_x\frac{\partial\psi}{\partial x}.$$

This, in turn, necessitates the presence of X and Z components of the electric field and current. However, the magnetic field has only a Y component

$$H_y = (\nabla \times A')_y = -i\gamma A_x - \frac{\partial A_z}{\partial x}.$$

From $\nabla \cdot A' = 0$, it follows that

$$\frac{\partial A_x}{\partial x} = i\gamma A_z.$$

Since

$$\frac{\partial A_x'}{\partial x} \sim \frac{1}{\lambda_o}A_x' \quad \text{and} \quad \frac{\partial\psi}{\partial x} \sim \frac{\kappa}{\lambda_n}\psi$$

(A' falls away from surface as e^{-r/λ_0} and ψ as $e^{-\kappa r/\lambda_0}$) ϕ can be shown to be approximately

$$\phi \sim \frac{2V^2\gamma\kappa}{c\omega} A_z \sim \frac{2V^2}{c^2} \kappa A_z \ll A_z \,.$$

The ratio of the A' and ϕ terms in Equation (21) is then of the order of $4\kappa^2(V^2/c^2)$. Since $V^2/c^2 \sim 10^{-5}$, the charge effects will become noticeable for materials with $\kappa > 50$, i.e., only for the extreme type-II superconductors. Since most of the materials fall outside this range of κ, we are justified in assuming the charge to be zero, and can use Equation (23) for the required solution. There are, however, some theoretical inconsistencies connected with this assumption. Zero charge implies that $\nabla \cdot A' = 0$ and $\nabla \cdot J = 0$. But

$$-\frac{4\pi}{c} \lambda_o^2 \nabla \cdot J = \psi^2 \nabla \cdot A' + A' \cdot \nabla \psi^2 = A'_x \frac{\partial \psi^2}{\partial x} = 0 \,.$$

Since $\partial\psi/\partial x \neq 0$, then $A_x' 0 = 0$. However, assuming $A_x = 0$ makes it impossible to satisfy $\nabla \cdot A' = 0$. As we have already seen, the X component of the vector potential is very small; hence the error in assuming the divergence of the current to be zero is of the same order as that in the assumption of zero charge.

After all these considerations, we are finally left with the equilibrium form of Ginzburg–Landau equations [Equation (23)] :

$$-\frac{\lambda_o^2}{\kappa^2} \frac{\partial^2\psi}{\partial x^2} + a^2\psi - \psi + \psi^3 = 0 \,. \tag{24}$$

$$\frac{\partial^2 a}{\partial x^2} + \left(\frac{\omega^2}{c^2} - \gamma^2 - \frac{\psi^2}{\lambda_o^2} \right) a \doteq \frac{\partial^2 a}{\partial x^2} - \frac{\psi^2}{\lambda_o^2} a = 0 \tag{25}$$

Unfortunately, there is no general solution to these equations. However, some special cases, such as that of very thin films ($\psi = $ constant) or of the extreme type-I material ($\kappa \sim 0$), can be readily solved. Another case for which a solution exists is that of $\kappa = \infty$. In this case, however, as we have seen, the charge effects are not negligible and must be taken into account.

PROPAGATION ALONG THIN FILMS

Ginzburg has shown[8] that when

$$
\left(\frac{\kappa d}{\lambda_o} \right)^2 \ll 1 ,
$$

where d is the film thickness, the field $h = H/(\sqrt{2}\, H_c)$ and the corresponding potential, a, are given by

$$
a_z = - \frac{h_o \cosh\left(\psi_o \dfrac{x}{\lambda_o} \right)}{\psi_o \sinh\left(\psi_o \dfrac{d}{\lambda_o} \right)} , \tag{26}
$$

$$
h_y = \frac{da_z}{dx} = \frac{h_o \sinh\left(\psi_o \dfrac{x}{\lambda_o} \right)}{\sinh\left(\psi_o \dfrac{d}{\lambda_o} \right)} , \tag{27}
$$

where $H_o = 2\pi I/c$ and I is the film current. The second relation between a and ψ_o is given by

$$
\psi_o^2(\psi_o^2 - 1) = - \frac{h_o^2 \left[1 + \dfrac{\sinh \dfrac{2\psi_o d}{\lambda_o}}{2\psi_o d/\lambda_o} \right]}{2 \sinh^2 \dfrac{\psi_o d}{\lambda_o}} . \tag{28}
$$

For $\psi_o d/\lambda_o \ll 1$, i.e., $d \ll \lambda_o$, this becomes

$$
h_o^2 = \psi_o^4(1 - \psi_o^2) \frac{d^2}{\lambda_o^2} .
$$

---

[3] V. L. Ginzburg, "Critical Current for Superconducting Films," *Doklady Ac. Sc. USSR*, (Translation), Vol. 3, p. 102, Jan.-Feb. 1958.

These relations are relatively independent of the G-L coupling constant, κ.

To determine the propagation constant, γ, we must also consider the wave propagating in the enclosing dielectric space. For the propagation to be possible, the waves in different media must have the same propagation constant. This is ensured by equating the tangential field components at the boundary.[9]

The wave equation in the dielectric medium reduces to

$$\frac{\partial^2 H_y}{\partial x^2} - k^2 H_y = 0 . \tag{29}$$

where

$$k^2 = \gamma^2 - \frac{\omega \epsilon^2}{c^2} .$$

Assuming the wave to die away for $x = \infty$ we have

$$H_y = H_o e^{-kx} \tag{30}$$

where we have already imposed the boundary conditions.

The electric field in the Z direction is given by

$$E_z = \frac{ick}{\epsilon \omega} H_o e^{-kx} \tag{31}$$

In the superconducting medium

$$E_z = -\frac{1}{c}\frac{\partial A'}{\partial t} = -\frac{i\omega}{c} A_z' , \tag{32}$$

and equating the two components at $x = 0$,

$$\frac{ck}{\omega} H_o = \frac{\omega}{c} A_z'(0) .$$

[9] J. C. Swihart, "Field Solution for a Thin-Film Superconducting Strip Transmission Line," *Jour. Appl. Phys.*, Vol. 32, p. 461, March 1961.

But

$$A_z(0) = \sqrt{2}H_c\lambda_o a_z(0) = -\frac{H_o\lambda_o(1 + \cosh \psi_o d)}{\psi_o \sinh\left(\psi_o \dfrac{d}{\lambda_o}\right)},$$

or

$$\frac{c^2}{\omega^2} - k = \frac{\lambda_o(1 + \cosh \psi_o d)}{\psi_o \sinh\left(\psi_o \dfrac{d}{\lambda_o}\right)} \doteq \frac{2\lambda_o^2}{\psi_o^2 d}$$

and

$$\gamma^2 = \frac{\omega^2}{c^2}\left(1 + \frac{\omega^2}{c^2}\frac{4\lambda_o^4}{\psi_o^4 d^2}\right). \tag{33}$$

Thus, the presence of the film will affect propagation only if $d \sim \lambda_o^2 \sim 10^{-10}$ cm. This condition, of course, can never be met in practice.

The reason for the negligible effect of the film on EM wave propagation is that the fields outside the superconductor are of the same order as the field in the film thickness (or in the penetration depth layer in the case of a bulk superconductor). The superconducting volume into which the fields penetrate is much smaller than that of the boundary dielectric space. The propagation constant, γ, is, therefore, very nearly that of the empty space.

One way in which γ can be appreciably affected by the presence of superconducting material is to ensure that the fields drop to a very small value at the boundary with the open space. Such a structure is a microstrip transmission line.

PROPAGATION CONSTANT OF A MICROSTRIP TRANSMISSION LINE

The microstrip transmission line, Figure 1, consists of a bulk type-I superconducting ground plane, a dielectric layer of thickness d_1, and a superconducting film of thickness d_2.

The solution for the vector potential in the ground plane is[2]

$$a_z = + h_o e^{-x/\lambda} \left\{ 1 + \frac{\kappa h_o^2}{\sqrt{2}(2-\kappa^2)} \left[\frac{\kappa}{4\sqrt{2}} e^{-2x/\lambda_0} - e^{-\sqrt{2}\kappa(x/\lambda_1)/[\kappa(\kappa+\sqrt{2})]} - \frac{3\kappa^3 + 3\sqrt{2}\,\kappa^2 - 8\kappa - 4\sqrt{2}}{4\sqrt{2}\,\kappa(\kappa+\sqrt{2})} \right] \right\} \quad (34)$$
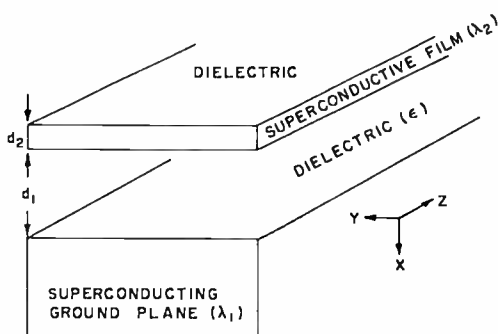
Fig. 1—Superconducting transmission line.

and

$$\psi = 1 - \frac{\kappa h_o^2}{\sqrt{2}(2-\kappa^2)} \left(e^{-\sqrt{2}\kappa x/\lambda_1} - \frac{\kappa}{\sqrt{2}} e^{-2x/\lambda_1} \right) \quad (35)$$

At the boundary with the dielectric layer $(x = 0)$

$$\psi_o = 1 - \frac{\kappa h_o^2}{2(\kappa + \sqrt{2})} \quad (36)$$

and

$$a_o = h_o \left[1 + \frac{\kappa h_o^2(\kappa + 2\sqrt{2})}{4(\kappa + \sqrt{2})} \right]. \quad (37)$$

Also,

$$E_{z_o} = -\frac{i\omega}{c}\sqrt{2}\,H_c\lambda_1 a_o = -\frac{i\omega\lambda_1}{c} H_o\alpha \ ,$$

where

$$\alpha = 1 + \frac{\kappa h_0^2 (\kappa + 2\sqrt{2})}{4(\kappa + \sqrt{2})}.$$

The solution for the wave equation in the dielectric and superconducting films has been outlined in the preceding section. By equating
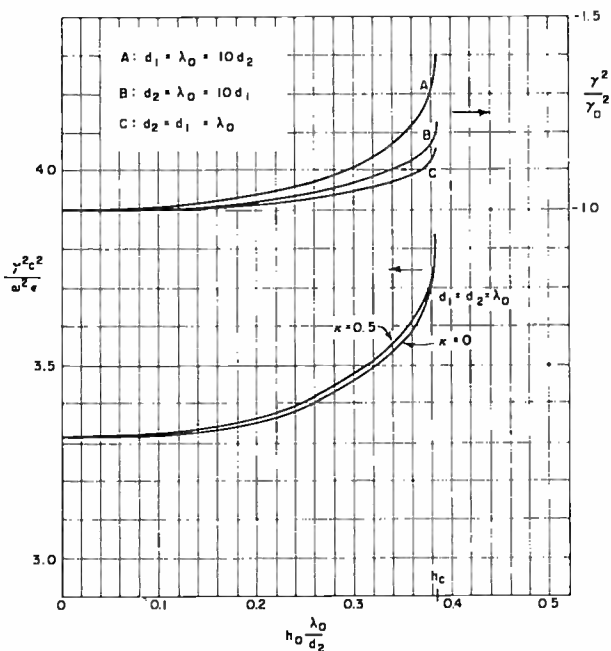
Fig. 2—Propagation constant as a function of magnetic field strength.

the tangential field components at the interfaces, all the arbitrary constants can be found, and we are left with one additional boundary condition which is used to determine the propagation constant, γ.

Omitting rather lengthy calculations, the propagation constant is finally given by

$$\gamma^2 = \frac{\omega^2 \epsilon}{c^2} \left(1 + \frac{\lambda_1}{d_1} \alpha + \frac{\lambda_2}{\psi^2 d_1} \coth \frac{\psi^2 d_2}{\lambda_2} \right), \tag{38}$$

where λ_1 and λ_2 are the zero field penetration depths in the ground

plane and the strip, respectively. Both α and ψ (the order parameter in the superconducting film) are field-strength dependent [Equations (37) and (28)].

Figure 2 shows the variation of γ with the magnetic field strength for a typical thin-film transmission line. The change in the propagation constant becomes appreciable as the magnetic field approaches the critical value for the film, given by

$$h_c \doteq \frac{2}{3\sqrt{3}} \frac{d_2}{\lambda_2}.$$

At the critical field the rate of change becomes infinite.

Curves A, B, and C in Figure 2 show the variation in the propagation constant, normalized to the zero-field value, for different line geometries. The change is greatest in Curve A, which is for a thin superconducting film line ($d_2 = \lambda_2/10$), and least where both the strip and the dielectric layer are fairly thick ($\sim \lambda$). The change in γ is more sensitive to the thickness of the strip than to that of the dielectric layer. In general, material with a larger zero-field penetration depth will have a larger variation of γ for a given line geometry.

CONCLUSIONS

We have shown that, to the extent that charge effects can be neglected, the equilibrium Ginzburg–Landau equations can be used together with Maxwell's equations to derive the solution of EM wave propagation in superconductors.

The transmission line structures that have negligible fields along the boundary with open space and have transverse dimensions of the order of penetration depth exhibit appreciable variation of the propagation constant with magnetic field (or current). This makes them suitable for distributed traveling-wave parametric amplifiers.

SELECTION DIVERSITY WITH
NON-ZERO CORRELATIONS

BY

A. SCHMIDT*

Summary—Frequency diversity is an attractive technique for communication in a fading environment. When correlation (between channels) is present, diversity performance will, in general, be poorer. This paper presents the results of a study, using a correlation model, to determine how performance varies for a fixed order of diversity (up to 8th order) as the amount of correlation varies. Curves are also given to illustrate the effect of crowding more and more diversity carriers into a fixed available bandwidth.

THIS PAPER is concerned with the power advantage to be obtained through the use of selection diversity in a Rayleigh fading environment when the diversity channels are correlated to some degree. The subject is of interest because previous analysis (assming zero correlation) has indicated that transmitter-switched frequency diversity is an attractive technique for communication in a fading medium.

Diversity performance will, in general, be poorer when correlation is present. The amount of correlation for any given configuration of diversity carriers will vary with physical conditions. The more diversity carriers that are packed into a fixed r-f band, however, the more correlated will be the fading at the carrier frequencies. Thus, two questions arise:

(1) For a fixed order of diversity, how does performance vary as the amount of correlation varies?

(2) For a fixed r-f bandwidth, how does performance vary as the order of diversity (and, therefore, the correlation between carriers) is increased?

The dual-diversity, non-zero-correlation case was analyzed by Staras;[1] apparently no work has been done on higher orders of selection diversity. Pierce and Stein[2] have written an extensive paper on

---

* Formerly with RCA Defense Communications Systems Division; now with Tracor, Inc., New York, N. Y.
[1] H. Staras, "Diversity Reception with Correlated Signals," *Jour. Appl. Phys.*, Vol. 27, p. 93, Jan. 1956.
[2] J. N. Pierce and S. Stein, "Multiple Diversity with Nonindependent Fading," *Proc. I.R.E.*, Vol. 48, p. 89, Jan. 1960.

arbitrary orders of diversity with non-zero correlation, but for maximal-ratio combining.

Some results of the present study for dual, triple, quadruple, 6th, and 8th-order diversity are shown in Figures 1 through 5. Note that surprisingly little advantage is lost with envelope correlation coefficients of adjacent channels as large as 0.7. (This is partly a consequence of the model assumed, which is described below.) Curves for zero correlation coefficient were not included, since they fall within 0.1 or 0.2 dB of the $\rho = 0.1$ curve.
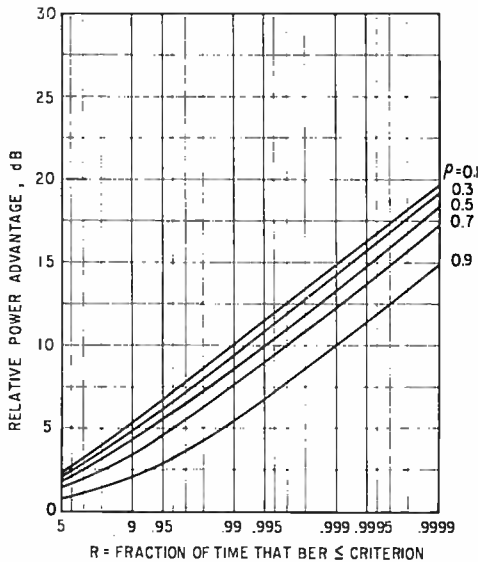
Fig. 1—Relative power advantage for selection diversity of order $N = 2$ for various values of envelope correlation coefficient, ρ.

The "relative power advantage" plotted in the figures is defined as the amount by which the average transmitted power can be reduced, relative to a non-diversity system, using the same modulation technique for the same R in both cases. R is the fraction of time that bit error rate (BER) is equal to or less than some preassigned criterion. (It is shown in the Appendix that the power advantage is a function of R, but is independent of the BER criterion.) It is assumed that the diversity switching is performed at the transmitter (no power splitting among the diversity channels).
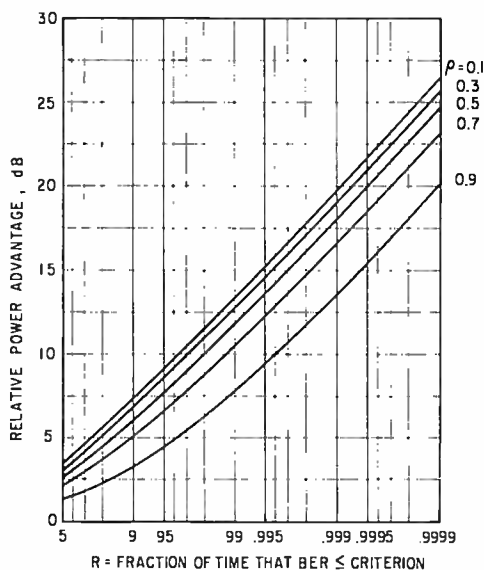
Fig. 2—Relative power advantage for selection diversity of order $N = 3$ for various values of envelope correlation coefficient, ρ.
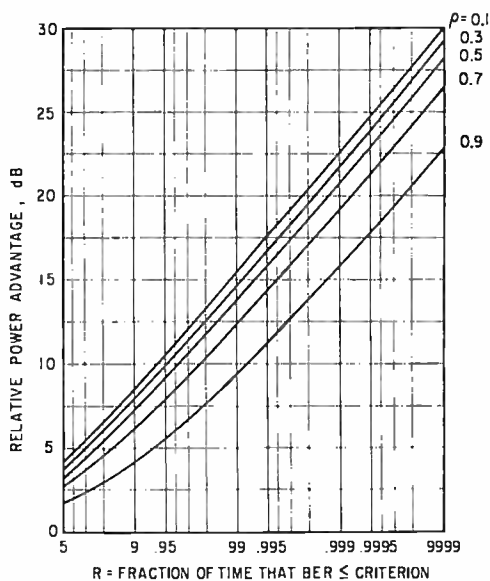
Fig. 3—Relative power advantage for selection diversity of order $N = 4$ for various values of envelope correlation coefficient, ρ.
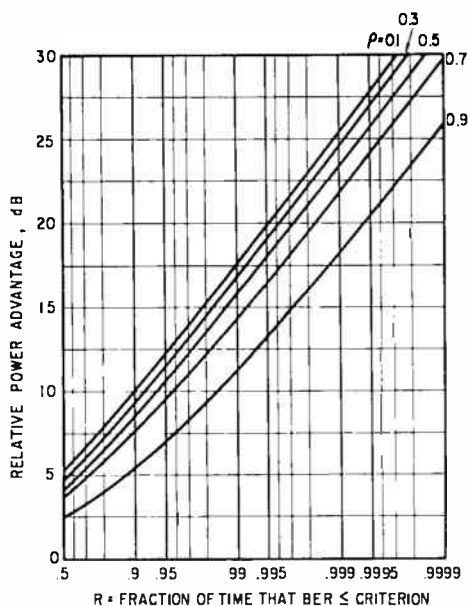
Fig. 4—Relative power advantage for selection diversity of order $N = 6$ for various values of envelope correlation coefficient, ρ.
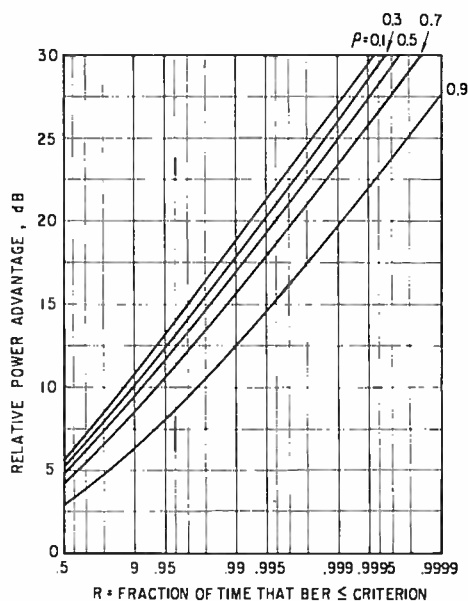

Fig. 5—Relative power advantage for selection diversity of order $N = 8$ for various values of envelope correlation coefficient, ρ.

Figure 6 illustrates the effect of crowding more and more diversity carriers into a fixed available r-f bandwidth. An envelope correlation coefficient of 0.3 was assumed between two carrier frequencies at the maximum separation permitted by the available bandwidth and the outer side bands. With this correlation coefficient fixed, additional diversity carriers were inserted between the two frequencies, with correlation coefficient between adjacent carriers determined by the model
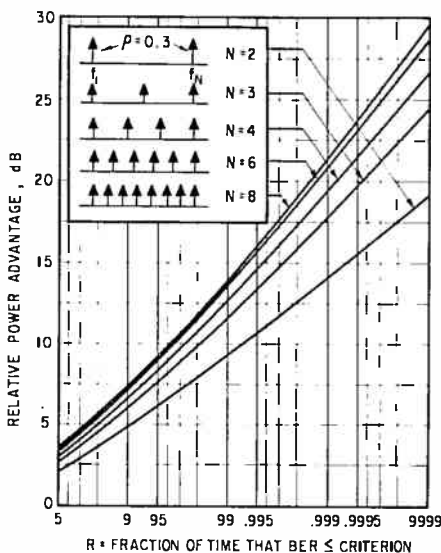
Fig. 6—Relative power advantage for various orders of diversity within a fixed r-f bandwidth $(\rho(f_N - f_1) = 0.3)$.

described below. Note that the power advantage (for a fixed R) increases more and more slowly as the order of diversity is increased. The results in Figure 6, when extrapolated to $N = 16$, indicate an additional advantage (over $N = 8$) of roughly 0.3 dB at $R = 0.9$, 1 dB at $R = 0.99$, 1.6 dB at $R = 0.999$, and 2.4 dB at $R = 0.9999$. These figures serve only to indicate approximate magnitude, as there is considerable uncertainty in the extrapolation.

The correlation model used requires some explanation here (a more detailed discussion is given in the Appendix). Assume N equally spaced diversity carriers $\{C_1, C_2, \ldots C_N\}$, and consider the following

matrix of correlation parameters (covariances):

$$
\begin{array}{cccccccc}
1 & \lambda & \lambda^2 & \lambda^3 & \cdot & \cdot & \cdot & \cdot & \lambda^{N-1} \\
\lambda & 1 & \lambda & \lambda^2 \\
\lambda^2 & \lambda & 1 & \lambda \\
\lambda^3 & \lambda^2 & \lambda & 1 \\
\cdot & & & & \cdot \\
\cdot & & & & & \cdot \\
\cdot & & & & & & \cdot \\
\cdot & & & & & & & \cdot \\
\lambda^{N-1} & & & & & & & & 1
\end{array}
$$

This matrix states that the correlation parameter for adjacent channels (C_1 and C_2, C_2 and C_3, etc.) is λ in each case. The correlation parameter for two carriers separated by two "increments" (e.g. C_1 and C_3) is λ^2; for two carriers separated by three increments, it is λ^3; etc. (If the mathematical model is applied to types of diversity other than frequency, the word "carrier" becomes "diversity channel", and "increment" is appropriately interpreted.)

This model implies that the correlation parameter between frequencies f_a and f_b is an exponential function of the frequency separation:

$$\text{corr. par.} = \exp\left(-c\,|f_a - f_b|\right) .$$

The experimental data examined is too sketchy and scattered to either support or refute this assumed law, but it seems reasonable. S. O. Rice[3] derived an expression of the form

$$\rho = \exp\left[-c(f_a - f_b)^2\right] .$$

However, this expression assumes that the density of the scattering cloud varies (spatially) according to a normal law. Rice cautions against taking the formula too seriously, especially away from the central region.

APPENDIX—ANALYSIS OF CORRELATED DIVERSITY

The first step is to find the joint probability density function (pdf) of the N correlated Rayleigh variables. The pdf is then integrated in N dimensions to obtain the probability that the largest of the N signals exceeds a level Z. The power advantage can be expressed in terms of Z and the corresponding probability.

---

[3] S. O. Rice, "Statistical Fluctuations of Radio Field Strength Far Beyond the Horizon," *Proc. I.R.E.*, Vol. 41, Feb. 1953.

For an arbitrary set of correlation parameters, the formula for the joint pdf appears intractable (see Reference (4), p. 40, Theorem 7). However, a useful formula exists in at least one special case. In Theorem 6 on p. 37 of Reference (4), let $n = 2$ and $p = N$. Then the theorem may be paraphrased as follows.

Assume N two-dimensional random vectors. Let the components of the kth vector, x_k and y_k, be Gaussian random variables with zero means. Let all the x_k's be independent of all the y_k's. Let the x_k's and y_k's have the same covariance matrix, M_N, such that its inverse, $W_N = M_N^{-1}$, is a Jacobi matrix (i.e., $w_{jk} = 0$ for $|j - k| > 1$). Let $r_k = \sqrt{x_k^2 + y_k^2}$. Then the joint pdf of the r_k's is given by

$$g_N(r_1, r_2, \ldots, r_N)$$

$$= |W_N| \exp\left\{ -\frac{1}{2} \sum_{i=1}^{N} w_{ii} r_i^2 \right\} \prod_{j=1}^{N} r_j \prod_{k=1}^{N-1} I_o(|w_{k,k+1}| r_k r_{k+1}) .$$

Since we are mainly interested in power, let $u_k = r_k^2/2$, and assume unit mean power for each of the r_k's (i.e., $\overline{u}_k = 1$). Then

$$f_N(u_1, u_2, \ldots, u_N)$$

$$= |W_N| \exp\left\{ -\sum_{i=1}^{N} w_{ii} u_i \right\} \prod_{j=1}^{N-1} I_o(|w_{j,j+1}| 2\sqrt{u_j u_{j+1}}) .$$

Among the matrices with Jacobi inverses is the Toeplitz matrix with elements

$$m_{jk} = \lambda^{|j-k|} ;$$

that is, the matrix is of the same form as the matrix of correlation parameters given earlier.

Here, the correlation parameter λ is the covariance of the x-components (or y-components) of two *adjacent* diversity carriers. It is the same as the parameter k used by Staras[1] and others,

$$k = \frac{\sqrt{(\overline{x_1 x_2})^2 + (\overline{x_1 y_2})^2}}{\overline{x_1^2}}$$

[4] K. S. Miller, "Multidimensional Gaussian Distributions," John Wiley and Sons, 1964.

except that here it is assumed that $\overline{x_1 y_2} = 0$ (this assumption is discussed later). The relation between λ and the correlation coefficient of the envelopes, ρ, is[5]

$$\rho = \frac{\overline{r_1 r_2} - \overline{r_1}\,\overline{r_2}}{[(\overline{r_1^2} - (\overline{r_1})^2)(\overline{r_2^2} - (\overline{r_2})^2)]^{1/2}}$$

$$= \frac{4[E(\lambda) - \left(\dfrac{1 - \lambda^2}{2}\right) K(\lambda)] - \pi}{4 - \pi}$$

where E and K are the complete elliptic integrals. It turns out that

$$\rho \approx \lambda^2 .$$

It can be shown that the elements of the inverse covariance matrix W_N are

$$w_{ij} = \begin{cases} \dfrac{1}{1 - \lambda^2}, & i = j = 1, N \\[2mm] \dfrac{1 + \lambda^2}{1 - \lambda^2}, & i = j = 2, 3, \ldots, N - 1 \\[2mm] -\dfrac{\lambda}{1 - \lambda^2}, & |i - j| = 1 \\[2mm] 0, & |i - j| > 1 \end{cases}$$

so that there are only three types of non-zero element, with values that are independent of N. The value of the determinant of W_N is

$$|W_N| = \frac{1}{(1 - \lambda^2)^{N-1}} .$$

The assumption that the x_k's (in-phase components) are independent of the y_k's (quadrature components) requires some discussion. The formula given above for $g_N(r_1, r_2, \ldots, r_N)$ agrees with the stand-

―――――――――――
[5] J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals," Vol. 24, p. 62 and p. 156, McGraw-Hill Book Co., New York, 1950.

ard formula in the case $N = 2$ (e.g., Equation 3.7-13 in Reference (6)); in the standard formula, however, $\overline{x_1 y_2}$ may assume non-zero values. The meaningful parameter appears to be the sum $(\overline{(x_1 x_2)}^2 + (\overline{x_1 y_2})^2)$, since the individual terms do not appear independently in the result. Thus it seems that for two fading carriers, the term $\overline{x_1 y_2}$ may be arbitrarily set equal to zero and all the correlation lumped into the $\overline{x_1 x_2}$ term (also equal to $\overline{y_1 y_2}$). This is supported by the following evaluation.

Since a Rayleigh fading carrier is statistically identical to narrow-band Gaussian noise, use Rice's approximate representation (Reference (6), Equation 2.8-6) for one of the fading carriers:

$$I_1(t) = \sum_{n=1}^{N} c_n \cos (\omega_n t - \phi_n) ,$$

where the c_n's are proportional to the power spectral density at the frequencies f_n,

$$c_n = \lfloor 2w(f_n)\Delta f \rfloor^{\frac{1}{2}}$$

and the ϕ_n's are independent random variables, each uniformly distributed over the interval $(0, 2\pi)$. Assuming the second fading carrier power spectrum has the same shape, translated to a different center frequency,

$$I_2(t) = \sum_{n=m}^{N+m-1} c_{n-m+1} \cos (\omega_n t - \theta_n) .$$

Assume the center frequencies are Ω_1 and Ω_2. I_1 and I_2 can be rewritten in the form

$$I_1(t) = x_1(t) \cos \Omega_1 t - y_1(t) \sin \Omega_1 t ,$$

$$I_2(t) = x_2(t) \cos \Omega_2 t - y_2(t) \sin \Omega_2 t ,$$

where

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \sum_{n=1}^{N} c_n \begin{bmatrix} \cos \\ \sin \end{bmatrix} \phi_n ,$$

---

[6] S. O. Rice, "Mathematical Analysis of Random Noise," *Bell Syst. Tech. Jour.*, Vol. 23, p. 282, July 1944 and Vol. 24, p. 46, Jan. 1945.

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \sum_{k=1}^{N} c_k \begin{bmatrix} \cos \\ \sin \end{bmatrix} \Theta_k \,,$$

and

$$\Phi_n = (\omega_n - \Omega_1)t - \phi_n \,,$$

$$\Theta_k = (\omega_{k+m} - \Omega_2)t - \theta_k$$

$$= (\omega_k - \Omega_1)t - \theta_k \,.$$

The $\overline{x_1 x_2}$ term is evaluated as

$$\overline{x_1 x_2} = \sum_{n=1}^{N} c_n^2 \, \overline{\cos \Phi_n \cos \Theta_n}$$

$$= \frac{1}{2} \sum_{n=1}^{N} c_n^2 \left\{ \overline{\cos \left[2(\omega_n - \Omega_1)t - (\phi_n + \theta_n)\right]} + \overline{\cos (\phi_n - \theta_n)} \right\} \,.$$

Since it appears unlikely that there is any "preferred" angle in nature, the sum angle $\phi_n + \theta_n$ must be uniformly distributed (there is no apparent reason for any bias) even if ϕ_n and θ_n are not independent. Hence

$$\overline{\cos \left[2(\omega_n - \Omega_1)t - (\phi_n + \theta_n)\right]} = 0$$

for any t, and

$$\overline{x_1 x_2} = \frac{1}{2} \sum_{n=1}^{N} c_n^2 \, \overline{\cos (\phi_n - \theta_n)}$$

and similarly for $\overline{y_1 y_2}$. For the $\overline{x_1 y_2}$ term,

$$\overline{x_1 y_2} = \sum_{n=1}^{N} c_n^2 \, \overline{\cos \Phi_n \sin \Theta_n}$$

$$= \frac{1}{2} \sum_{n=1}^{N} c_n^2 \, \overline{\sin (\phi_n - \theta_n)}$$

by the same argument as above, and similarly

$$\overline{x_2 y_1} = \frac{1}{2} \sum_{n=1}^{N} c_n^2 \overline{\sin (\theta_n - \phi_n)} = - \overline{x_1 y_2} .$$

In the absence of evidence to the contrary, the most reasonable assumption is that the difference angle $(\phi_n - \theta_n)$ is *symmetrically* distributed about zero, so that $\overline{\sin (\phi_n - \theta_n)} = 0$ for each term in the sum, and

$$\overline{x_1 y_2} = \overline{x_2 y_1} = 0 .$$

The two carriers just considered could have been any of the N involved in the joint pdf of the envelopes. Thus any x_j is independent of any y_k, $j \neq k$ (of course this is also true for $j = k$, as usual). Miller's theorem is therefore applicable.

The probability that the instantaneous power of the largest of N carriers exceeds the level Z is

$R = 1 -$ pbty {all N carriers $< Z$}

$$= 1 - \int_{u_1=0}^{Z} \cdot \cdot \cdot \int_{u_N=0}^{Z} f_N(u_1, u_2, \ldots, u_N) du_1 du_2 \ldots du_N .$$

Since all the u_k's have been normalized to the unit mean power, this is actually the probability that the largest carrier power will exceed Z times its mean level.

For given λ and N, R can be calculated (by computer) as a function of Z, thus establishing Z as a function of R. The relative power advantage can be found as follows.

For the reference (non-diversity) system, the bit-error rate (BER) is a function of the signal (or signal/noise) power level:

$$BER_1 = F_1(s_1)$$

and the required level s_{10} to meet a BER criterion p_o (maximum acceptable BER) is

$$s_{10} = F_1^{-1}(p_o) .$$

Regarding the level at any time as a random variable S_1, the probability (or fraction of time) that this performance criterion will be met is

$$R_1 = \text{pbty } \{S_1 > s_{10}\} = G_1\left(\frac{s_{10}}{m_1}\right)$$

where m_1 is the mean of S_1.

Correspondingly, for the diversity system under evaluation,

$$\text{BER}_2 = F_2(s_2) .$$

For the same BER criterion p_o,

$$s_{20} = F_2{}^{-1}(p_o), \text{ and}$$

$$R_2 = \text{pbty } \{S_2 > s_{20}\} = G_2\left(\frac{s_{20}}{m_2}\right).$$

The two systems are to be compared on the basis of the mean signal levels (m_1 and m_2) required to achieve identical probabilities of achieving the required BER, p_o. Thus set

$$R_1 = R_2 = R .$$

The inverse relations

$$\frac{s_{10}}{m_1} = G_1{}^{-1}(R)$$

$$\frac{s_{20}}{m_2} = G_2{}^{-1}(R)$$

lead to the relative power advantage

$$\text{Adv } (N,\lambda,p_o,R) = \frac{m_1}{m_2}$$

$$= \frac{s_{10}}{s_{20}} \frac{G_2{}^{-1}(R)}{G_1{}^{-1}(R)}$$

$$= \frac{F_1^{-1}(p_o)}{F_2^{-1}(p_o)} \frac{G_2^{-1}(R)}{G_1^{-1}(R)} .$$

In many cases this power advantage is independent of the BER criterion p_o, either because the modulation techniques of the reference and diversity systems are identical (so that $F_1 = F_2$) or because the signal/noise ratio requirements as functions of BER have a constant-dB difference (so that $F_1^{-1}/F_2^{-1} = $ constant, as with binary differential phase-shift keying and frequency-shift keying, for example). Assume that both systems use the same modulation technique, so that $s_{10} = s_{20}$, and

$$\mathrm{Adv} = \frac{G_2^{-1}(R)}{G_1^{-1}(R)} .$$

For the non-diversity reference system, the probability that the power of the Rayleigh fading carrier will exceed the level s_1 is

$$R_1 = G_1 \left(\frac{s_1}{m_1} \right) = \exp\left(-\frac{s_1}{m_1} \right)$$

so that

$$\frac{s_{10}}{m_1} = G_1^{-1}(R) = -\log R .$$

For the diversity system, the normalized parameter $Z = s_2/m_2$ cannot be written explicitly as a formula, but is known numerically as a result of calculating $R_2(Z)$,

$$Z = G_2^{-1}(R) .$$

Therefore

$$\mathrm{Adv} = \frac{m_1}{m_2} = \frac{G_2^{-1}(R)}{G_1^{-1}(R)} = \frac{G_2^{-1}(R)}{-\log R} .$$

This is the quantity plotted in Figures 1 through 6.

ADAPTIVE DETECTION MODE WITH THRESHOLD CONTROL AS A FUNCTION OF SPATIALLY SAMPLED CLUTTER-LEVEL ESTIMATES

BY

H. M. FINN AND R. S. JOHNSON

RCA Defense Electronic Products
Camden, N. J.

Summary—This paper describes automatic surveillance modes for the detection of targets embedded in statistically nonstationary and nonuniform extended-clutter target environments. These modes yield a false-alarm probability invariant to changes in the clutter level, and a detection efficiency free of the loss associated with the predetection hard-limiting approach to false-alarm rate regulation. The false-alarm-probability regulation mechanism involves making the detection threshold proportional to a spatially sampled maximum-likelihood estimate of the output variance of a cell under test where this variance is due to the clutter limiting environment. The weights employed implicitly relate the a priori information of the target-scattering function of the clutter. The analysis treats the nonfluctuating target and the Swerling fluctuating target models. Detection efficiency is noted to be a monotonically increasing function of the number of resolution-cell outputs employed in the clutter-level-estimation procedure. For the computed cases, where 100 resolution cells are employed, the detection efficiency is comparable to that of an ideal detection situation where the noise level is known and fixed thresholds can be employed. Mismatch conditions are analyzed, and the procedure is shown to be relatively insensitive to errors made in a priori assumptions of the shape factor of the clutter target-scattering function. Target pair resolution capabilities of these adaptive detection modes are also analyzed. The introduction of a second target in one of the threshold control cells introduces a masking effect equivalent to a 1-dB loss in detection efficiency for a worst-case analysis where 100 resolution cells are employed in the threshold-control system.

I—INTRODUCTION

AN AUTOMATIC detection mode for a surveillance radar is one in which some function of the envelope detector output at a resolution cell is compared with a threshold number, and the target is declared present if the threshold is exceeded. The mode is contrasted, e.g., with the case where detection is effected solely by an operator employing manual controls, and video displays. The false-alarm probability when operating in the automatic detection mode in a clutter environment is extremely sensitive to small changes in the average value of the back-scattered energy from the clutter targets assuming that no special provisions have been introduced to avoid this sensitivity.

This sensitivity is clearly viewed in Figure 1. As can be seen, for the case where the threshold is originally set for a false-alarm probability of 10^{-8}, a 3-dB increase in the total noise power density (clutter plus thermal noise) yields a 10,000-fold increase in the false-alarm probability—an increase that typically would significantly exceed the radar data-handling rate.
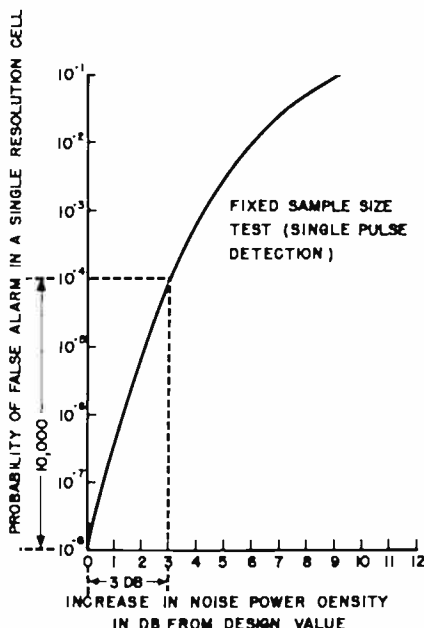

Fig. 1—False alarm probability for fixed threshold detector.

Since most clutter environments are characterized as yielding unknown and time-varying average values of back-scattering energy (statistically nonstationary returns), automatic target detection in a clutter environment cannot be achieved with a conventional receiver configuration and a fixed threshold. In the approach considered here, false-alarm sensitivity is reduced by making the threshold proportional to a maximum-likelihood estimate of the magnitude of a target scattering function that represents the extended clutter target phenomena.[1-3] This

[1] H. M. Finn, "Adaptive Detection with Regulated Error Probabilities," *RCA Review*, Vol. 29, p. 653, Dec. 1967.

[2] H. M. Finn, "Adaptive Radar Detection with Regulated Error Probabilities in Extended Clutter Target Environments," *Proc. 5th Ann. Allerton Conf. on Circuit and System Theory*, Univ. Illinois, Oct. 1967.

[3] H. M. Finn, "Adaptive Detection in Clutter," *Proc. Nat. Electronics Conf.*, Vol. XXII, p. 562, 1966.

estimate of the level of the clutter-target back-scattering energy can be made with a single transmission. It is based on the outputs of the radar resolution cells surrounding the cell under test, cells that are also assumed to be covered by the clutter cloud.

The shape of the extended clutter target scattering function (i.e., the law describing the change in clutter level as a function of range and doppler) is assumed to be known. The magnitude of this scattering function is unknown, however, and is the quantity estimated in the spatial sampling estimation procedure. This procedure yields a false-alarm probability invariant to changes in the clutter level if the stochastic properties of the encountered clutter are in agreement with the assumptions implicit in the threshold control procedure. The detection efficiency for a radar detection mode based on this procedure is also shown to be a monotonically increasing function of the number of resolution cells employed in the threshold-control formulation. For the cases considered where 100 resolution cells are employed in the clutter-level-estimation procedure, the detection efficiency is comparable to that of an ideal detection situation where the noise power density is a known quantity and a fixed threshold can be employed. This result is interesting, since it indicates that the false-alarm-rate regulation procedure described is free of the loss of detection efficiency usually associated with predetection hard-limiting approaches to false-alarm-rate regulation.

Analyses are made both for the matched case, where the basic extended clutter target model encountered by the radar is in agreement with the assumptions employed in the adaptive threshold control function, and also for the mismatched case, where some of the stochastic properties of the encountered clutter differ from the assumed model. The target pair resolution capabilities of this adaptive detection mode are also analyzed.

II—BASIC CONCEPT

The basic concept of the adaptive threshold-control detection mode developed in this paper can be considered a generalization of a threshold control procedure employed for false-alarm-rate regulation in an additive thermal noise environment free of spatially distributed clutter targets. This conventional system, the constant false-alarm-rate regulation system (CFAR) was designed to adapt to changes in the variance of the detector output when the variance in the noise power density is due either to changes in the additive thermal noise level or to the receiver sensitivity. In a typical CFAR system of this type, the square-law-detected receiver output is sampled during the radar dead time

(say, a time interval that is equivalent to 100 resolution cells, or $100/B$, where B is the processing system bandwidth) and the threshold is made proportional to the sum of these outputs. This sum can be shown to be proportional to the maximum-likelihood estimate (MLE) of the variance of the noise output. Essentially, the process involves the following steps: (1) the noise source is sampled (for this conventional thermal noise limiting system, the radar dead time is employed for
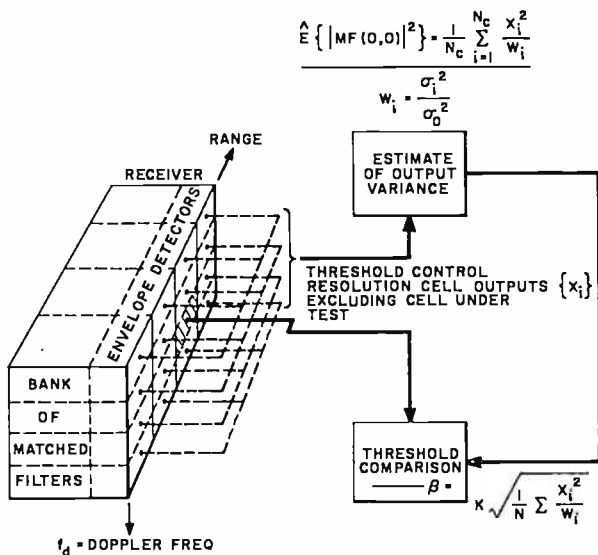

Fig. 2—Concept of automatic threshold control procedure.

the sampling interval), (2) a maximum-likelihood estimate is made of the variance of the noise output (based on a sufficient number of statistically independent noise samples), and (3) the threshold is made proportional to this estimate of the variance of the output. For the case where a linearly detected output is employed as the decision statistic, the threshold is made proportional to the square root of this estimate.

A logical extension or generalization of this procedure for the case where the noise is the sum of the receiver noise and the back-scattering from the spatially distributed clutter targets involves the following steps (see Figure 2). After a transmission, the outputs of the resolution cells that surround the cell under test and are also covered by the extended-clutter target cloud are sampled. These samples are used to obtain a maximum-likelihood estimate of the variance of the output of the cell under test $\hat{E}\{|MF(\tau_o,f_o)|^2\}$. The sampled linearly detected out-

put of the resolution cell under test is then compared with a detection threshold D_o that is made proportional to the square root of this estimate of the output variance:

$$D_o = K\sqrt{\hat{E}\{|MF(\tau_o,f_o)|^2\}}.$$

The statistical hypothesis test for the presence or absence of the signal return from a point target embedded in the extended-clutter target and additive thermal noise environment is assumed here to be performed at each beam position as a fixed-sample-size test. A target-present hypothesis is accepted if at least one of the outputs of the N_F resolution cells is greater than its threshold, i.e., if $\bigcup_{i=1}^{N_F} (X_i > D_i)$ is true.

A multiple resolution-cell case is assumed for this problem. That is to say, the hypothesized target may appear in an admissible-target parameter space where, for example, a range-doppler parameter space is considered and a number of contiguously spaced matched filters cover the admissable parameter space. The resolution-cell dimensions are the inverse of the transmitted-signal bandwidth (range), and the doppler dimension is the inverse of the signal time duration. An extension of this adaptive approach to multiple-stage decision processors and a development of detection modes that regulate both sets of error probabilities (false alarm and false dismissal) are included in References (1) and (2).

III—DEVELOPMENT OF SPATIALLY SAMPLED MAXIMUM LIKELIHOOD ESTIMATE OF RECEIVED CELL OUTPUT VARIANCE

The maximum-likelihood estimate of the ensemble average of the square of the matched filter output of the cell under test $E\{|MF(\tau_o,f_o)|^2\}$ is for some important cases of interest

$$\hat{E}\{|MF(\tau_o,f_o)|^2\} = \frac{1}{N}\sum_{i=1}^{N} \frac{X_i^2}{W_i}. \tag{1}$$

This estimate of the output variance is a weighted sum of the N outputs of the resolution cells surrounding the cell under test (also covered by the same extended clutter target phenomenon). $X_i = |MF(\tau_i,f_i)|$ represents the sampled matched-filter output of the ith resolution cell (centered about the range delay coordinate of τ_i, and the doppler shift of f_i). The set of weights W_i are real numbers implicitly relating the a priori information. The following assumptions are made in the development of Equation (1).

(1) The extended-clutter-target phenomenon is assumed to be of the noncoherent type, i.e., clutter returns from disjoint portions of the parameter space are uncorrelated. Precipitation and chaff clutter phenomena are examples of noncoherent clutter.

(2) The clutter target scattering function $\rho(\tau,f)d\tau df$ may be represented as the product of an unknown magnitude C (to be estimated) and a known shape factor $\rho'(\tau,f)d\tau df$.

$\rho(\tau,f)d\tau df$ represents the average value of the return energy from the clutter targets in the region with range delay between τ and $\tau + d\tau$, and a doppler shift between f and $f + df$. As stated previously, the shape of the target scattering function $\rho'(\tau,f)d\tau df$ is assumed to be a priori information. The set of weights (W_i) in Equation (1) is shown to be a function of this shape factor. The magnitude, C, of the clutter level is the parameter for which a maximum likelihood estimate is to be made.

(3) Finally, either or both of the following cases is assumed: (a) the contribution of the clutter return far exceeds the additive thermal noise contribution to the output variance; (b) the clutter level (the target-scattering function) is uniform over the parameter space involved in the estimation procedure.

The development Equation (1) begins with the expression of the output variance of the ith resolution cell for the case of noncoherent clutter:

$$E\{|MF(\tau_i,f_i)|^2\} = \int \int |X(\tau,f)|^2 \rho(\tau + \tau_i, f + f_i)\, d\tau df + N_o . \qquad (2)$$

N_o is the additive thermal noise power density of the receiver, and the component of variance due to the incoherent clutter cloud is expressed in Equation (2) as the two-dimensional convolution of the Woodward's signal-ambiguity function[4] $|X(\tau,f)|^2$ and the target-scattering function of the extended clutter phenomenon $\rho(\tau,f)d\tau df$. (For example, see Westerfield et al[5], Price and Green[6], Gersch[7], and Van Trees[8].

[4] P. M. Woodward, *Probability and Information Theory, with Applications to Radar*, McGraw-Hill, New York, 1953.

[5] E. C. Westerfield, R. H. Prager, and J. L. Stewart, "Processing Gains Against Reverberation (Clutter) Using Matched Filters," *IRS Trans. Inf. Theory*, (Matched Filter Issue) Vol. IT-6, p. 342, June 1960.

[6] R. Price and P. E. Green, Jr., "Signal Processing in Radar Astronomy-Communication Via Fluctuating Multipath Media," Lincoln Lab. Tech. Report 234, Oct. 1960.

[7] W. Gersch, "On Processing the Radar Return from an Ionized Trail," Purdue Univ. Report 5, Sec. 1, 1964.

[8] H. L. Van Trees, "Optimum Signal Design and Processing for Reverberation-Limited Environments," *IEEE Trans. Military Electronics*, Vol. MIL-9, p. 212, July/Oct. 1965.

The maximum-likelihood estimator $\hat{E}\{|MF(\tau_o,f_o)|^2\}$ is developed to be[1]

$$\hat{E}\{|MF(\tau_o,f_o)|^2\} = \frac{1}{N}\sum_{i=1}^{N}\frac{X_i^2}{W_i} \tag{3}$$

where $W_i = \gamma_i^2/\gamma_o^2$ and the set $\{\gamma_i\}$ are defined by

$$\gamma_i^2 = \int\int |X(\tau,f)|^2\rho'(\tau + \tau_i, f + f_i)\,d\tau df$$

That is to say

$$W_i = \frac{\displaystyle\int\int |X(\tau,f)|^2\rho'(\tau + \tau_i, f + f_i)\,d\tau df}{\displaystyle\int\int |X(\tau,f)|^2\rho'(\tau + \tau_o, f + f_o)\,d\tau df}, \tag{4}$$

and it is noted that W_i is a function of known quantities, the signal ambiguity function $|X(T,f)|^2$, and the shape factor of the target scattering function $\rho'(\tau,f)\,d\tau df$.

For the case where the target scattering function is relatively constant over the resolution cells involved in the estimation procedure so that $\gamma_i^2 = \gamma_j^2$ for all i and j, $W_i = 1$ for all i and

$$\hat{E}\{|MF(t_o,f_o)|^2\} = \frac{1}{N}\sum_{i=1}^{N}X_i^2. \tag{5}$$

The threshold D_o is now made proportional to the square root of this output variance estimate, or

$$D_o = K\sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{X_i^2}{W_i}}. \tag{6}$$

and the resultant false-alarm probability is shown in Section VI to be simply a function of the threshold control constant K, and the number of resolution cells N. Consequently it is invarient to changes in the clutter level. In addition, the detection efficiency is shown to be relatively good. It is noted in this section and in Section V that the detection efficiency is a monotonically increasing function of the number of

resolution cells employed in the control function, and that for the computed cases where 100 resolution cells were employed in this estimation procedure, the detection efficiency is comparable to that of the ideal detection situation where the noise levels are known quantities and need not be estimated so that fixed thresholds can be employed.

Consequently, this adaptive detection procedure of making the detection threshold proportional to a spatially sampled maximum-likelihood estimate of the output variance due to the extended-clutter target
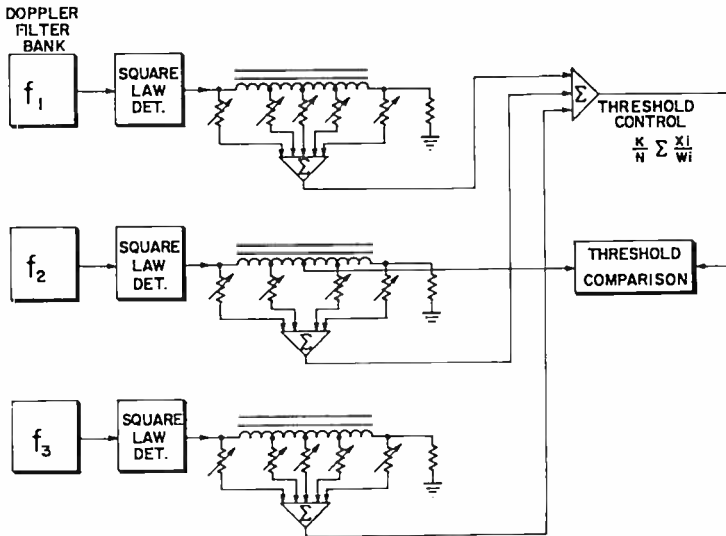
Fig. 3—Implementation for automatic threshold control system.

phenomena and thermal noise is not only an intuitively satisfying procedure but it is shown to yield a constant false-alarm rate with good detection efficiency.

Figure 3 shows the implementation of this procedure with tapped delay lines. For the pulse doppler radar assumed in this diagram, the outputs of the range-doppler cells surrounding the cell under test are effectively sampled by employing separate delay lines for each doppler channel involved. The taps are placed at delay intervals of $1/B$ (B is the bandwidth of the transmitted signal), and resistors at the output of the taps and feeding the summing amplifiers provide the weights $\{W_i\}$. Square-law detectors are employed at the output of each doppler filter. The delay line center tap at the output of one of the doppler channels is designated in this diagram as the cell under test. A separate threshold determination of the type indicated in this diagram

would be required for each doppler channel, because the weighted sum employed as a threshold must exclude the cell under test.

IV—PERFORMANCE EVALUATION OF AUTOMATIC THRESHOLD CONTROL PROCEDURE

Typical sets of computed performance data of this automatic-threshold-control procedure are shown in Figures 4 through 16. The first set
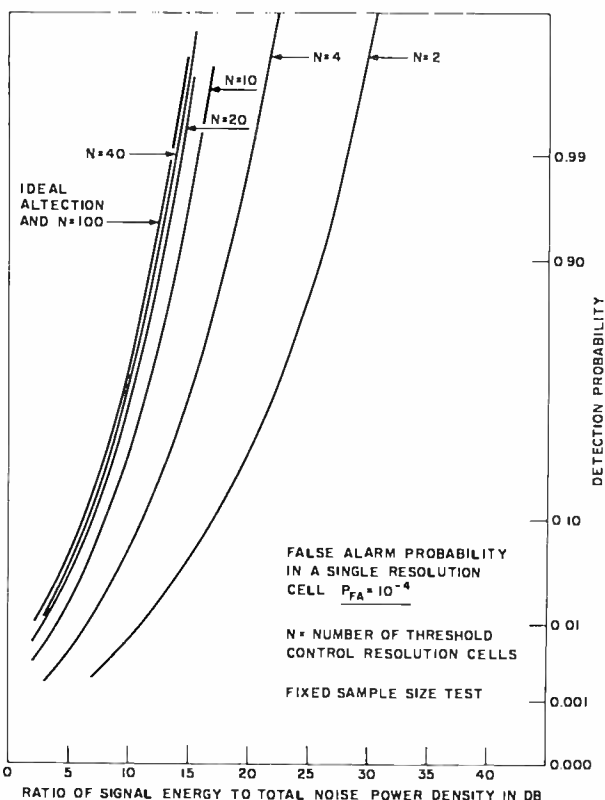
Fig. 4—Automatic threshold control detection probability for nonfluctuating target, $P_{FA} = 10^{-4}$.

of data (Figures 4 through 12) treat the matched case, i.e., the case where the encountered statistics of the back-scattering from the clutter region agree with those that have been assumed; the first-order-probability density function of the clutter return is Rayleigh distributed, and the shape of the encountered target scattering function is that assumed in setting the weights $\{W_i\}$ in the threshold control equation (Equation (6)).

Expressions for detection and false-alarm probability are developed later in this paper and used to compute the detection characteristics for typical sets of surveillance conditions given in Figures 4 through 11. Both the nonfluctuating and the Swerling target types have been considered. In the figures, the detection probability is given as a function of the ratio of average signal energy to total noise power density
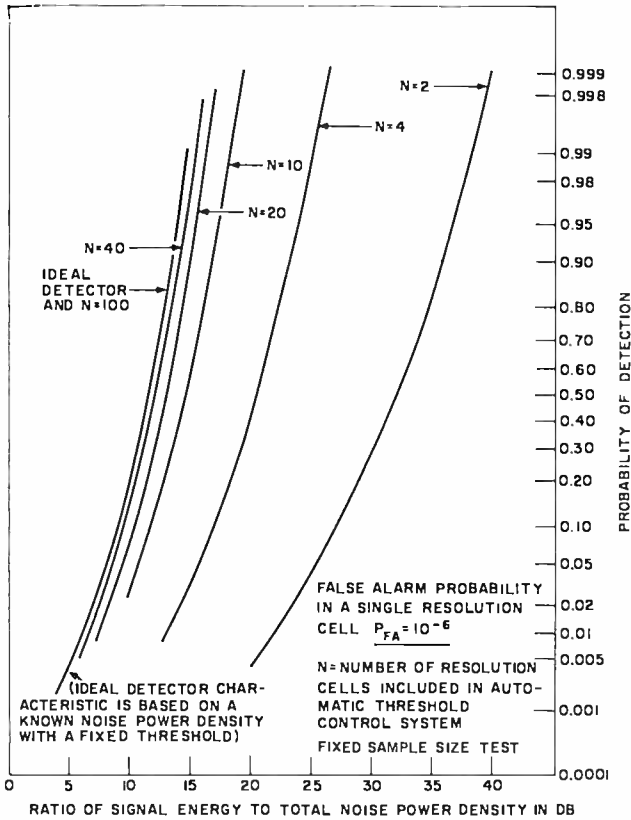
Fig. 5—Automatic threshold control detection probability for nonfluctuating target, $P_{FA} = 10^{-6}$.

(thermal plus clutter noise) for the case where the threshold factor K has been set for a specified single-cell detection probability. Values of false-alarm probability ranging from 10^{-4} to 10^{-10} have been considered. From the expression for the single-cell false-alarm probability (Equation (24)), it is clear that for a given number N of resolution cells involved in threshold control, the false-alarm probability is a function of K and is independent of the noise-power density. Thus, this is a CFAR system.

N is a fixed parameter in Figures 4 through 11. As N increases, the detection process becomes more efficient, until the detection characteristic approaches that of the ideal detection situation. The characteristic curve for the case of 100 resolution cells essentially coincides with the detection characteristic curve for the ideal detection situation.

The results of the detection computations are summarized in Figure 12, where loss in dB is given as a function of N. This loss represents
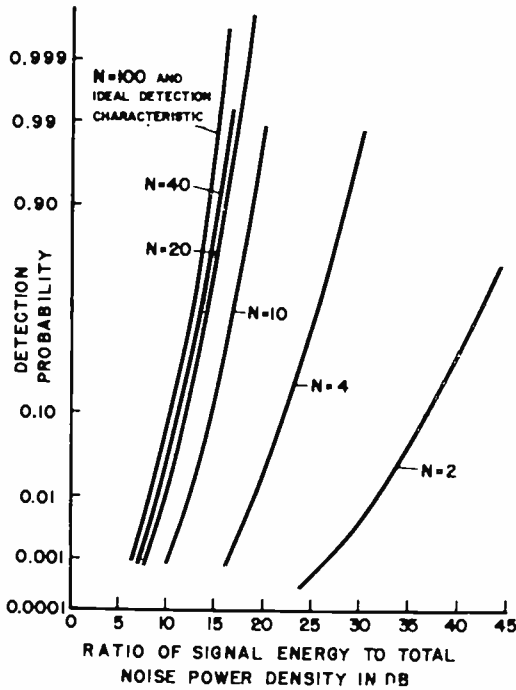
Fig. 6—Automatic threshold control detection probability for nonfluctuating target, $P_{FA} = 10^{-8}$.

the increase in the ratio of signal energy to noise power density required for the automatic threshold detection mode to achieve a prescribed detection and false-alarm probability. The standard of comparison is the ideal situation of a detection mode employing a fixed threshold and operating with a known noise-power density. It can be seen that as the number of resolution cells employed in the measurement of the noise power density increases, the loss decreases and asymptotically approaches zero. For 30 threshold control resolution cells, the loss is typically less than 1 dB. For a given number of resolution cells, the loss varies inversely with the false-alarm probability requirement.

Data computed for the performance of the threshold-control procedure when mismatch conditions exist are summarized in Figures 13 through 16. The edge-effect performance analysis results are presented in Figures 13-15. These performance data are developed in detail later for the case where some of the threshold control resolution cells are outside the clutter cloud and 'see' only the additive thermal noise of the receiver, and the remainder of the control cells receive the back-
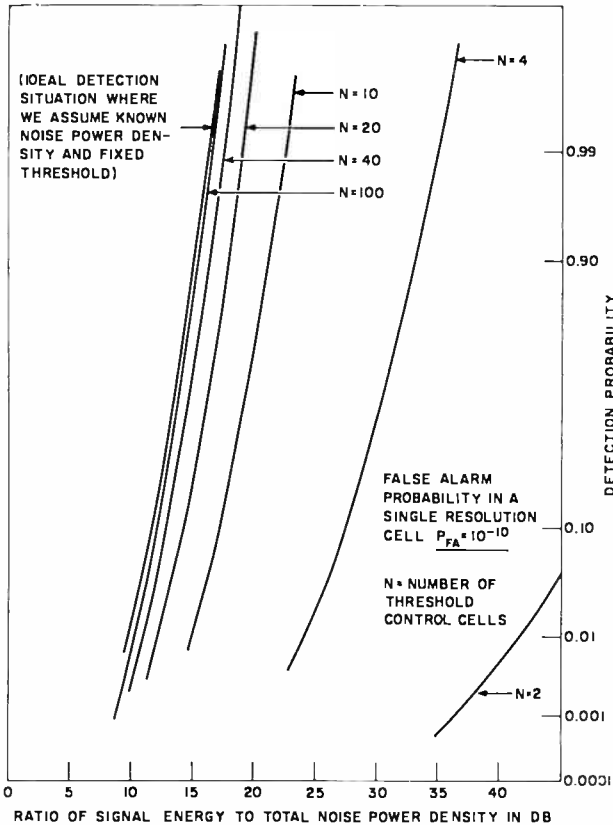
Fig. 7—Automatic threshold control detection probability for nonfluctuating target, $P_{F.1} = 10^{-10}$.

scattering from the clutter targets. Step-function discontinuities have been assumed to achieve the most conservative evaluation.

Another type of mismatch is the case where the ground-return clutter has been assumed to have a first-order probability density function agreeing with the Rayleigh distribution, but where the back-scattering contains a coherent as well as incoherent components and the first-order

probability distribution adheres to the Rice distribution rather than the Rayleigh. The results of an analysis of this type of mismatch are given in Figure 16.

Last, an analysis is given of the threshold control procedure for the case where a pair of targets are embedded in the extended clutter
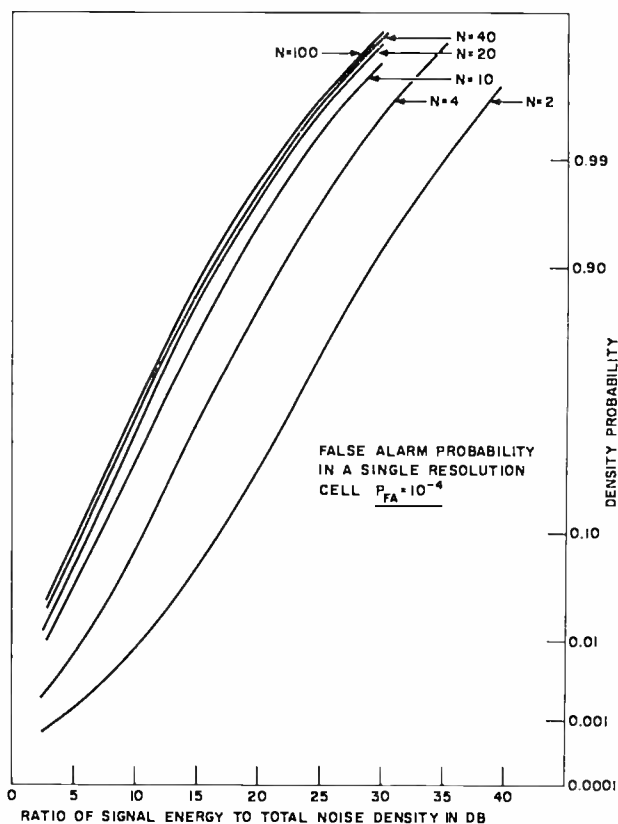
Fig. 8—Automatic threshold control detection probability for Swerling Case #3 target, $P_{FA} = 10^{-4}$.

target 'cloud'. The family of computed curves in Figure 18 (see Section XI) present the detection probability for the case where one of the targets is in the cell under test, and the other one appears in one of the threshold control cells. As can be seen in Figure 18, this loss in detection efficiency is a monotonically decreasing function of the number of resolution cells. For the case of a detection probability of 0.5, single-cell false alarm probability of 10^{-8}, and 100 resolution cells employed in the threshold control, the loss in detectability due to the

presence of this second target is approximately 1 dB for the worst-case situation considered in the analysis.

V—PROBABILITY DENSITY FUNCTION OF THRESHOLD

Since the threshold D_o in Equation (6) is a function of the N random variables $\{X_i\}$, D_o is also a random variable. In the assumed
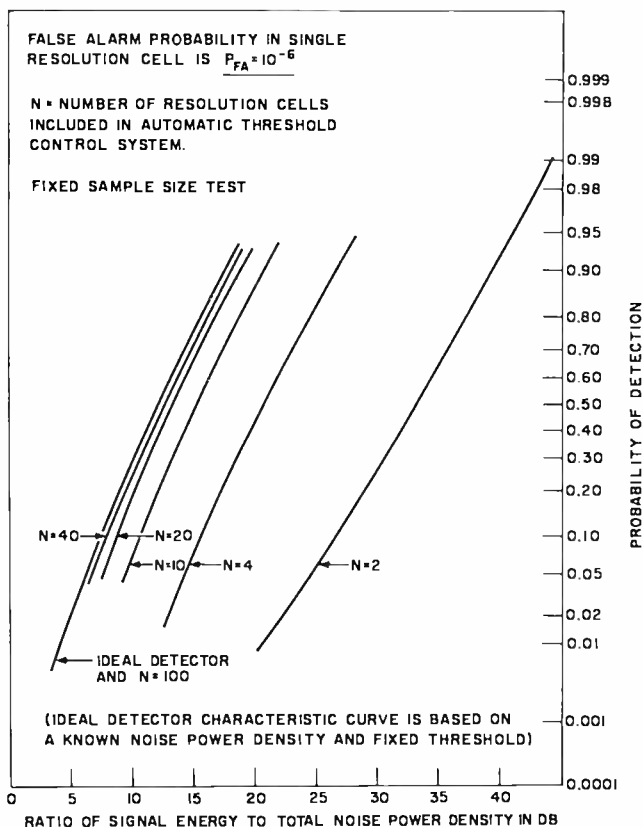
Fig. 9—Automatic threshold control detection probability for Swerling Case #3 target, $P_{FA} = 10^{-6}$.

case the $\{X_i\}$ are statistically independent and Rayleigh distributed, i.e.,

$$f(X_i)\,dX_i = \frac{X_i}{\sigma_i^2} \exp\left\{ -\frac{X_i^2}{2\sigma_i^2} \right\}\, dX_i$$

where

$$\sigma_i^2 = (E\{|MF(\tau,f_i)|^2\})/2 . \tag{7}$$

Then, the probability density function of the threshold D_0 is shown to be

$$f(D_o)dD_o = \frac{2}{(N-1)!\left(K\sigma_o\sqrt{\dfrac{2}{N}}\right)^{2N}} D_o^{2N-1} \exp\left(-\left(\frac{D_o}{K\sigma_o\sqrt{\dfrac{2}{N}}}\right)^2\right)dD_o$$

$$(8)$$
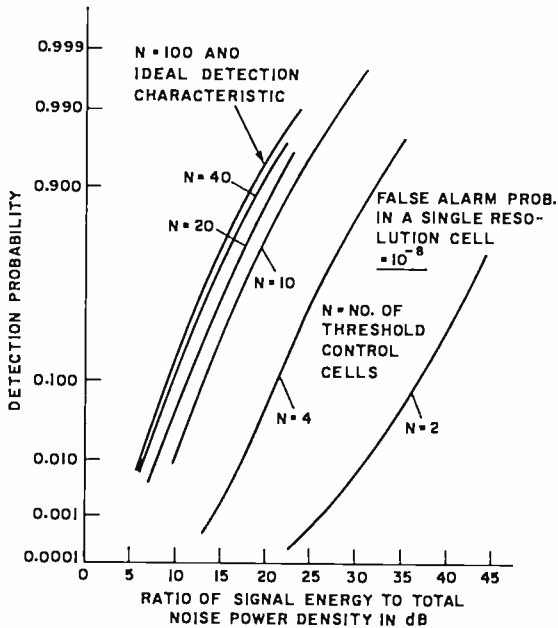
Fig. 10—Automatic threshold control detection probability for Swerling Case #3 target, $P_{FA} = 10^{-8}$.

As can be seen, the threshold probability density function is a function of the threshold parameter K, the number of resolution cells involved in threshold control N, and σ_o. The final expression for the false-alarm probability is shown in the following section to be independent of σ_o.

VI—DETECTION PROBABILITY WITH AUTOMATIC THRESHOLD CONTROL— NONFLUCTUATING TARGET AND SWERLING CASE #1 AND #3

The expression for the detection probability of a fixed sample size test incorporating the threshold control procedure is developed in
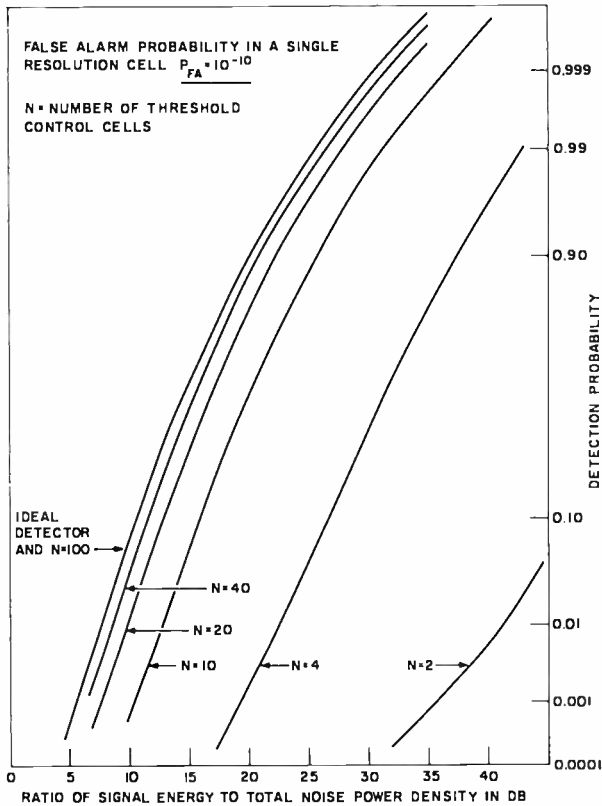
Fig. 11—Automatic threshold control detection probability for Swerling Case #3 target, $P_{FA} = 10^{-10}$.
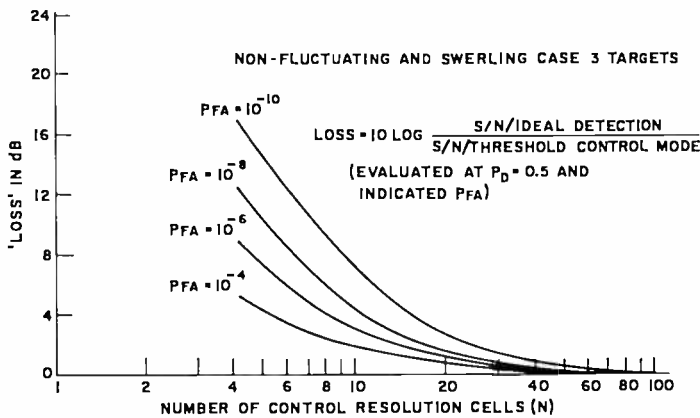
Fig. 12—Loss of adaptive detection mode versus number of threshold control resolution cells, N.
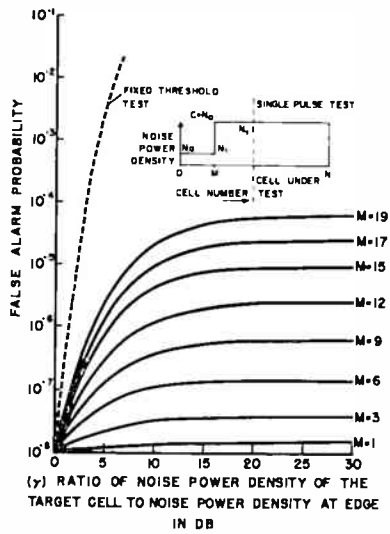
Fig. 13—False alarm probability versus γ where $\gamma > 1$.
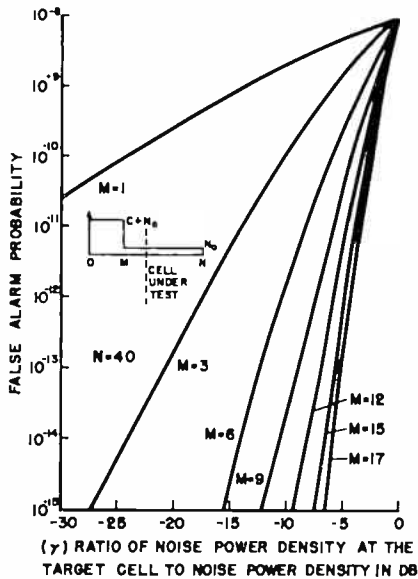
Fig. 14—False alarm probability versus γ where $\gamma < 1$.
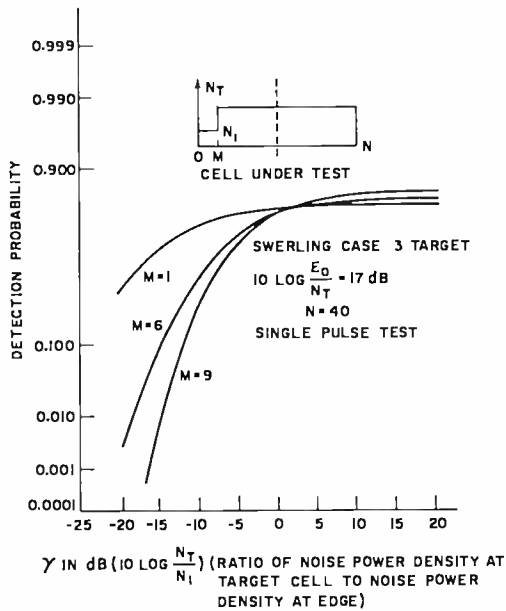
Fig. 15—Probability of detection versus γ.
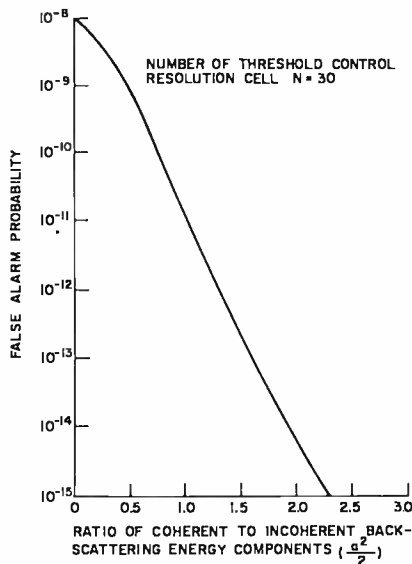
Fig. 16—False alarm probability versus ratio of coherent to noncoherent back-scattering energy components.

Appendix I for the nonfluctuating target, and in Appendix III for the Swerling Case #3 target type.

The single-pulse probability of detection of a nonfluctuating target based on comparing the linearly detected matched filter output X with the threshold D_o, can be expressed as

$$P_D = \int_{D_o=0}^{\infty} \int_{X=D_o}^{\infty} \frac{X}{N_T/2} \exp\left\{\frac{-X^2-E}{N_T}\right\} I_o\left(\frac{\sqrt{EX}}{N_T\cdot2}\right) dX f(D_o) dD_o , \qquad (9)$$

where $f(D_o)dD_o$ is the threshold density function defined in Equation (8). Since the domain of D_o include all positive real numbers, the integration with respect to D_o is from 0 to infinity.

The portion of the integrand in Equation (9) that is expressed as

$$f(X) dx = \frac{X}{N_T/2} \exp\left\{\frac{-X^2-E}{N_T}\right\} I_o\left(\frac{\sqrt{EX}}{N_T/2}\right) dX \qquad (10)$$

is the conditional probability density function of the sampled linearly detected matched-filter output X when E is the input signal energy, and N_T is the total (clutter plus thermal noise) power density;[9,10] $N_T = 2\sigma_o^2$. Introducing the change of variable $Z = X/\sqrt{N_T/2}$ in Equation (9), and defining $\alpha = \sqrt{2E/N_T}$, Equation (9) becomes

$$P_D = \int_{\beta=0}^{\infty} Q(\alpha,\beta) f(\beta;K,N) d\beta , \qquad (11)$$

where $Q(\alpha,\beta)$ is the well-known Q function[9,11] and is defined as

$$Q(\alpha,\beta) = \int_{z}^{\infty} Z \exp\left\{\frac{-Z^2-\alpha^2}{2}\right\} I_o(\alpha Z) dZ \qquad (12)$$

[9] C. W. Helstrom, *Statistical Theory of Signal Detection*, Pergamon Press, New York, 1960.

[10] W. W. Peterson, T. G. Birdsall, and W. C. Fox, "The Theory of Signal Detectability," *IRE Trans. Inf. Theory*, Vol. PGIT-4, p. 171, Sept. 1954.

[11] J. I. Marcum, *Table of Q Functions*, Rand Corp. RM 339, 1950.

and β is the normalized threshold, $\beta = D_o/\sqrt{N_T/2}$. The probability density function of β is

$$f(\beta;K,N)d\beta = \frac{2}{(N-1)!\left(K\sqrt{\dfrac{2}{N}}\right)^{2N}}\,\beta^{2N-1}\exp\left\{-\left(\frac{\beta}{K\sqrt{\dfrac{2}{N}}}\right)^2\right\}d\beta\,. \tag{13}$$

The solution of Equation (11) is shown in Appendix I to be

$$P_D = 1 - \frac{\gamma^2}{\gamma^2+1}\sum_{m=0}^{N-1}\frac{\exp\{-\alpha^2/(\gamma^2+2)\}}{m!}\left(\frac{2}{\gamma^2+2}\right)^n L_m(\xi) \tag{14}$$

where $\gamma = K\sqrt{2/N}$ and $\xi = \alpha^2\gamma^2/[2(\gamma^2+2)]$ and $\{L_m(\xi)\}$ are Laguerre polynominals with the properties:

$$L_o(\xi) = 1, L_1(\xi) = 1 + \xi, L_{m+1}(\xi) = (\xi + 2m + 1)L_m(\xi) - m^2 L_{m-1}(\xi)\,. \tag{15}$$

For the Swerling case #3 target, the expression for the detection probability is

$$P_D = \int_{\beta=0}^{\infty}\int_{\alpha=0}^{\infty} Q(\alpha,\beta)f(\beta;K,N)\,d\beta f(\alpha;X)\,d\alpha\,, \tag{16}$$

where

$$f(\alpha;X)\,d\alpha = \frac{2\alpha^3}{\bar{X}^2}\exp\left\{-\frac{\alpha^2}{\bar{X}}\right\}d\alpha\,. \tag{17}$$

Performing the integration first with respect to α,

$$P_D = \int_{\beta=0}^{\infty}\left[1 + \frac{\beta^2\bar{X}}{(\bar{X}+2)^2}\right]\exp\left\{-\frac{\beta^2}{\bar{X}+2}\right\}f(\beta;K,N)\,d\beta\,. \tag{18}$$

The expression

$$P_D' = \left[1 + \frac{\beta^2 \bar{X}}{(\bar{X} + 2)^2} \right] \exp \left\{ - \frac{\beta^2}{\bar{X} + 2} \right\}. \tag{19}$$

in the integrand of Equation (18) represents the probability of detection of a Swerling Case #3 target when the average signal-energy-to-noise-power density is $\bar{X}$ and the threshold is β. Since β in our threshold control procedure is a random variable, however, we must perform the threshold integration in Equation (18).

In Appendix III, it is shown that Equation (18) reduces to

$$P_D = \left(\frac{\bar{X} + 2}{\bar{X} + 2 + 2 \dfrac{K^2}{N}} \right)^N \left[1 + \left[\frac{2K^2 \bar{X}}{(\bar{X} + 2)\left(\bar{X} + 2 + \dfrac{2K}{N} \right)} \right] \right]. \tag{20}$$

It is interesting to note that as $N \to \infty$, Equation (21) becomes

$$P_D = \left[1 + \frac{2K^2 \bar{X}}{(\bar{X} + 2)^2} \right] \exp \left\{ \frac{-2K^2}{\bar{X} + 2} \right\}. \tag{21}$$

If we replace $2K^2$ by β^2 in Equation (21), we obtain the same result as Equation (19) which is, in effect, the expression for the detection probability when the noise power density is known and the threshold can be fixed. This asymptotic relationship is in agreement with both intuition and the computed results. In effect, this relationship points out that when the number of resolution cell outputs, N, employed in the estimate of the noise power density becomes large enough, the variance of the estimate becomes negligibly small and the detection situation reverts to that of the ideal case where the noise-power density is known.

The detection probability for the Swerling Case #1 target is developed in the same manner by performing the integrations of Equation (16), but where now the density function of α is $f(\alpha) \, d\alpha = (\alpha/\bar{X}) \exp \left(-\alpha^2/2\bar{X} \right) d\alpha$. The result is

$$P_D = \left[\frac{1}{1 + K^2/[N(1 + \bar{X})]} \right]^N. \tag{22}$$

Here again it is noted that in the limit as $N \to \infty$, the detection probability reverts to the form of the probability of detecting a Case #1 target with a fixed threshold and a known noise power density of P_D $= \exp\ (-\beta^2/2(1 + \overline{X})$, where $\beta^2 = 2K^2$.

VII—DETECTION PROBABILITY FOR THE SWERLING CASES #2 AND #4 TARGET TYPES WITH AUTOMATIC THRESHOLD CONTROL

We will now develop the analytical expressions of the detection probability for the cases where J transmissions are employed, and the target cross section of the fluctuating target model is statistically independent from one transmission to another (the Swerling Cases #2 and #4 target types for example[12]). The sum of the J square-law-detector outputs of the zeroth cell

$$Z_o = \sum_{k=1}^{J} Y_{ok}$$

is compared with a threshold T_o made proportional to the estimate of the output variance. This estimate is based on the J outputs of each of the N resolution cells involved in the estimate and covered by the clutter cloud. The relationship of the threshold τ_o and D_o of Section II is $\tau_o = D_o/2$. $\tau_o = k_J \hat{E}\{|MF\{\tau_o,f_o\}|^2\}$, where $k_J = K_J^2/2$ (K is defined by Equation (16)) and

$$\hat{E}\{|MF(\tau_o,f_o)^2|\} = \frac{2}{JN} \sum_{i=1}^{N} \sum_{k=1}^{J} \frac{Y_{ik}}{W_i}. \tag{23}$$

The random variable Y_{ik} is the sample of square-law-detected output of the ith resolution cell after the kth transmission. The detection probability for this case may be expressed

$$P_D = \int_{T_o=0}^{\infty} P\{Z_o > T_o; \overline{X}, J\} f(T_o, J, N) dT_o. \tag{24}$$

---

[12] P. Swerling, "Probability of Detection for Fluctuating Targets," *IRE Trans. Inf. Theory* (Special Monograph Issue), Vol. IT-6, p. 269, April 1960.

Introducing the changes of variable

$$\xi_{ik} = \frac{Y_{ik}}{\sigma_o^2}, \quad z = \frac{Z_o}{\sigma_o^2}, \text{ and } \beta' = \frac{T_o}{\sigma_o^2},$$

the detection probability is

$$P_D = \int_{\beta'=0}^{\infty} P\{z > \beta';\overline{X},J\}f(\beta',J,N)\,d\beta' \qquad (25)$$

The probability density function of the threshold

$$\beta' = \frac{k_J}{JN} \sum_{i=1}^{N} \sum_{k=1}^{J} \frac{\xi_{ik}}{W_i}$$

is shown to be

$$f(\beta',J,N)\,d\beta' = \frac{\left(\dfrac{\beta'}{C}\right)^{JN-1}}{(JN-1)!} \exp\left\{\frac{\beta'}{C}\right\} \frac{d\beta'}{C}, \qquad (26)$$

where $C = k_J/JN$.

The conditional probability of detection based on the hypothesis that the threshold value is β' and the received average signal-energy-to-noise-power density is $\overline{X}$ has been derived elsewhere[1] for the Swerling Case #2 and #4 target models in the following forms.

For the Swerling Case #2 target model where the received signal to noise ratio (X) is distributed as

$$f(X;\overline{X})dX = \frac{1}{\overline{X}} \exp\left\{\frac{-X}{\overline{X}}\right\} dX \qquad (27)$$

$$P\{z > \beta';\overline{X},J\} = e^{-\alpha\beta'} \sum_{k=0}^{J-1} \frac{(\alpha\beta')^K}{K!} \qquad (28)$$

where

$$\alpha = \frac{1}{1 + \overline{X}} \qquad (29)$$

For the Swerling Case #4 target model, where the received signal-energy-to-total-noise-power density (X) is distributed as

$$f(X,\overline{X})\,dX = \frac{4X}{\overline{X}^2}\exp\left\{\frac{-2X}{\overline{X}}\right\}dX\,, \tag{30}$$

$$P\{z > \beta';\overline{X},J\} = e^{-\alpha'\beta'}\sum_{n=0}^{J}\frac{J!\left(\dfrac{1-\alpha'}{\alpha'}\right)^{n}}{n!\,(J-n)!}\sum_{K=0}^{J+n-1}\frac{(\alpha'\beta')^{K}}{K!} \tag{31}$$

where $\alpha' = 1/[1 + (\overline{X}/2)]$. $\tag{32}$

The analytical expressions for the Swerling Case #2 and #4 detection probabilities are then developed by performing the integration in Equation (25) employing Equations (26), (28), and (31). For the Swerling Case #2 target,

$$PD = \sum_{K=0}^{J-1}\frac{\left(\dfrac{\alpha\tau_J}{M}\right)^{K}(M+K-1)!}{\left(\dfrac{1+\alpha\tau_J}{M}\right)^{M+K}K!(M-1)!}. \tag{33}$$

For the Swerling Case #4 target,

$$P_D = \alpha^J\sum_{n=0}^{J}\frac{J!\left(\dfrac{1-\alpha'}{\alpha'}\right)^{n}}{n!\,(J-n)!}\sum_{K=0}^{J+n-1}\frac{\left(\dfrac{\alpha'\tau_J}{M}\right)^{K}(M+K-1)!}{\left(\dfrac{1+\alpha'\tau_J}{M}\right)^{M+K}K!(M-1)!}. \tag{34}$$

For both cases, $\tau_J = 2k_J$ and $M = NJ$, and α and α' are as previously defined for the Case #2 target and for the Case #4 target, respectively.

VIII—FALSE-ALARM PROBABILITY WITH AUTOMATIC THRESHOLD CONTROL

The single-cell false-alarm probability for the case where a single transmission is employed and the linearly detected matched-filter out-

put is used as the decision statistic is obtained from Equation (12) by setting $\alpha = 0$:

$$P_{FA} = \int_{\beta=0}^{\infty} \int_{z=\beta}^{\infty} Z \exp\left\{\frac{-Z^2}{2}\right\} dZ f(\beta;K,N) d\beta . \tag{35}$$

Performing the indicated integrations yields (see Appendix III)

$$P_{FA} = \left[\frac{1}{1+\dfrac{K^2}{N}}\right]^{N} \tag{36}$$

where it is seen that the false-alarm probability is independent of the noise power density. It is interesting to note that in the limit as $N \to \infty$,

$$P_{FA} = \exp\left\{-K^2\right\}, \tag{37}$$

and if we replace K^2 by $\beta^2/2$, Equation (37) is recognized as the expression for the false-alarm probability with a fixed threshold and known noise power density.

The false-alarm probability of a single resolution cell for the case where the sum of J square-law-detector outputs of the resolution cell are used as the decision statistic may be obtained from Equation (33) by setting $\overline{X} = 0$:

$$P_{FA} = \sum_{K=0}^{J-1} \frac{\left(\dfrac{\tau_J}{M}\right)^K (M+K-1)!}{\left(1+\dfrac{\tau_J}{M}\right)^{M+K} K!(M-1)!} \tag{38}$$

where $\tau_J = 2k_J$, and $M = NJ$. The expression of the false-alarm probability in Equation (35) is also noted to be independent of the clutter statistics.

The total false-alarm probability at the beam position under test may be symbolically represented as:

$$P_{FO} = P\left\{\bigcup_{i=1}^{N_F}\left(z_i > K\sqrt{\hat{E}\{|MF(\tau_i,f_i)|^2\}}\right)\right\} \tag{39}$$

where N_F is the number of resolution cells under test. This false-alarm probability is not obtained in the usual manner by the expression

$$P_{FO}' = 1 - (1 - P_{FA})^{N_F} \tag{40}$$

because the single-cell false-alarm probabilities for this threshold-control case are not independent of one another, i.e., some common outputs are employed in the output variance estimates. However, the above expression can be shown to be a good approximation to the total false-alarm probability, and for the case where $P_{FA} \ll 1$, a practical approximation is

$$P_{FO} \approx N_F P_{FA} . \tag{41}$$

IX—RELATED THRESHOLD-CONTROL PROCEDURE

The normalized periodogram detector[13,14] is an alternative threshold-control procedure for regulating the false-alarm probability in a clutter environment. In this CFAR system, a number of transmissions are used to detect a target in clutter, and the time interval between transmissions is such that statistically independent samples of the back scattering from the clutter region are obtained in a single resolution cell.

At the resolution cell undergoing test, the coherent sum of the N transmissions is envelope detected and is compared with a threshold made proportional to the sum of the square-law-detected outputs at that cell. The test is shown by Kelly, Reed, and Root[13] to be an implementation of the generalized likelihood ratio as the decision statistic for performing the alternate hypothesis test when the noise power density is an unknown but constant quantity for the duration of the N pulses, and statistically independent noise samples are obtained for each of the N transmissions. Probability distributions for this test are developed by Wishner.[14]

The normalized periodogram detector is somewhat restricted in its application in a clutter environment since it requires that the clutter return from each of the N transmissions in a common resolution cell be statistically independent and statistically stationary over the interval of the N transmissions. The adaptive detection mode in this paper requires only a single transmission.

[13] E. J. Kelly, I. S. Reed, and W. L. Root, "The Detection of Radar Echoes in Noise." Part I, *Jour. Soc. Indust. Appl. Math*, Vol. 8, p. 309, June 1960.

[14] R. P. Wishner, "Distribution of the Normalized Periodogram Detector," *IRE Trans. Inf. Theory*, Vol. IT-8, p. 342, Oct. 1962.

A basic philosophical difference between these two threshold-control approaches is that the adaptive detection procedure described here involves more of a pattern-recognition function, in that use is made of the a priori information relating to the *spatial* distribution of the extended-clutter target phenomena. The estimate is based on the samples of the surrounding resolution cells that are also covered by
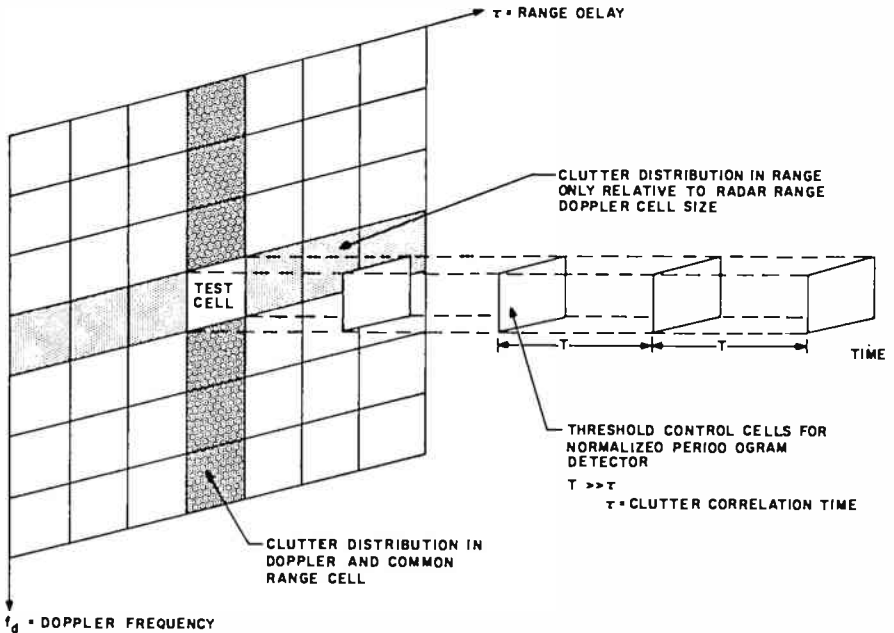

Fig. 17—Threshold control cells for particular clutter distributions.

the clutter cloud. The normalized periodogram detector, on the other hand, estimates the clutter-power density by taking successive samples in time of the clutter return in a common spatial resolution cell, the cell under test. Figure 17 is a schematic diagram of the relationship of these procedures.

X—MISMATCH ANALYSIS FOR ADAPTIVE DETECTION

The two basic mismatch cases considered are (1) the case where the shape of the encountered target scattering function $\rho(\tau,f)$ differs from that governing the threshold-control procedure and (2) the case where the first-order probability density function of the back scattering from clutter differs from that employed in the development of the maximum-likelihood estimate of the output variance. In both cases, we

want to know the effect of these conditions on the performance of the adaptive detection procedure.

Mismatch of the Shape of the Target-Scattering Function $p(\tau,f)d\tau df$

The general expressions for the probability of detection of a Swerling Case #3 target and the single-cell false-alarm probability for the case where the target scattering function employed differs from the encountered target-scattering function are developed in Appendix VI. These error probability expressions are presented below and are noted to be a function of the set $\{\alpha_i\}$ where α_i is the ratio of the assumed variance of the output of the ith cell to the actually encountered variance (i.e., $\alpha_i = \sigma_i^2/\underline{\sigma_i^2}$, where the sub-bar indicates the 'true' value.)

The Swerling Case #3 detection probability for this shape-mismatch case is

$$P_{D} = \frac{N}{K^2} \sum_{J=1}^{N} \frac{\eta_J}{\gamma_J} \left[1 + \frac{\overline{X}}{(X+2)^2\gamma_J} \right] \tag{42}$$

where

$$\gamma_J = \left[\frac{2}{\overline{X}+2} + \frac{\alpha_J N}{K^2} \right]. \quad \text{and} \tag{43}$$

$$\eta_J = \frac{\prod_{i=1}^{N} \alpha_i}{\prod_{i \neq J} (\alpha_i - \alpha_J)}. \tag{44}$$

The false-alarm probability in a single resolution cell is

$$P_{F,1} = \frac{N}{K^2} \sum_{J=1}^{N} \frac{\eta_J}{\left(1 + N\dfrac{\alpha_J}{K^2} \right)} \tag{45}$$

A reasonable approach for determining the sensitivity of the adaptive detection is to assume a particular variation of the clutter return with range, R, (e.g., R^{-3} for ground clutter and R^{-2} for precipitation and chaff), to adjust threshold-control weights in accordance with the

range variation affect assumed, and then to determine the performance when no clutter appears. A reversal of this mismatch situation could also be investigated. The weights could be made equal in accordance with an assumption of a uniform noise distribution in range, and then clutter distributed with an R^{-2} or R^{-3} relationship is encountered. For the case where clutter of an R^{-P} relationship is assumed but where a uniform distribution in range is actually encountered, the ratio of the assumed variance of total noise to the actual variance at the ith cell may be expressed as

$$\alpha_i = \left(\frac{R_o}{R_o + i\Delta} \right)^P = \left(\frac{1}{1 + i\dfrac{\Delta}{R_o}} \right)^P, \tag{46}$$

where Δ is the range resolution (for a 10-MHz bandwidth signal, $\Delta = 15$ meters), R_o is the range of the resolution cell under test (for a ground-clutter performance investigation, R_o could be say 30 km) and i is the range-cell index.

The ratio $\Delta/R_o = 5 \times 10^{-4}$ for these selected values; for the practical case where the total number of cells involved in the threshold control procedure is $N = 40$ (so that the index i extends from ± 1 to ± 20), it is clear that the set $\{\alpha_i\}$ of mismatch factors do not deviate significantly from unity. The developed expressions for P_D and P_{FA} also become less meaningful, since both are dependent on the products of very small numbers, i.e., $\prod_{i \neq j} (\alpha_i - \alpha_j)$ (see Equation (44).)

One argues then, that for a typical high-resolution radar in which a 10-MHz bandwidth signal is employed for operation in a clutter environment, a uniform weighting of the resolution-cell output surrounding the cell under test can be employed without significant degradation in the performance of the adaptive detection procedure. Another conclusion is that the type of shape mismatch analysis that would be meaningful involves consideration of edge affects. Edge effects occur when some of the resolution cells employed in the threshold-control procedure are in the clutter cloud while others are not and therefore 'see' only the receiver thermal noise.

Edge Effect Analysis

To analyze edge effects, we will consider the most severe condition, where the total noise power density as a function of range may be represented as a step function (see Figures 13, 14 and 15). The gen-

eral expressions for the detection and false-alarm probabilities of the adaptive-detection procedure when such discontinuities in the clutter distribution in range are encountered are developed in Appendix VI, where detection of a Swerling Case #3 target is considered. The detection probability is

$$P_D = \left[\frac{\gamma}{\alpha + \gamma}\right]^M \frac{\left[1 + \beta\left(\dfrac{M}{\alpha + \gamma} + \dfrac{N - M}{1 + \alpha}\right)\right]}{(1 + \alpha)^{N - M}}$$

where

$$\beta = \frac{\dfrac{2K^2}{N}\overline{X}}{(\overline{X} + 2)^2} \qquad \text{and} \quad \alpha = \frac{2K^2}{N(\overline{X} + 2)}. \tag{47}$$

The false-alarm probability in a single resolution cell is

$$P_{FA} = \frac{\left[\dfrac{\gamma}{\dfrac{K^2}{N} + \gamma}\right]^M}{\left(1 + \dfrac{K^2}{N}\right)^{N - M}} \tag{48}$$

where γ is the ratio of the total noise power density at the cell under test to that at the edge and the discontinuity in noise-power density occurs after the Mth resolution cell from one end. The sketch in Figure 13 should aid in a visualization of these parameters.

The affect of the edge discontinuity has been computed using Equations (47) and (48), and typical families of performance data are given in Figures 13, 14, and 15. The effects on the false-alarm probability regulation properties of the adaptive-detection procedure are given in Figures 13 and 14. In Figure 13, the cell under test is immersed in the clutter return, but a group of the resolution cells at the edge are not. As expected, for the severe discontinuities introduced, the false-alarm probability is no longer a constant but increases with increases in the noise-power-density discontinuity. However, even for the worst case

examined, the case where approximately one-half of the cells are in the clutter cloud and the other half are not, a 3-dB change in the noise-power density causes an increase in the false-alarm probability by a factor of less than 40, as compared to a 10^4 increase for the case of a fixed-threshold system. Figure 14 represents the reverse edge-effect condition. Here the cell under test is in the clear region, but a group of cells at the edge are immersed in the clutter cloud. A masking effect results. The threshold is raised unnecessarily due to the contributions of the cells in the cloud; consequently, the false-alarm probability is reduced. The masking effect is noted in Figure 15, where the probability of detection is plotted as a function of the ratio of the noise-power densities at the discontinuity, and where the signal energy to total noise-power density at the target cell is a fixed parameter. The detection probability is noted to fall for the case where the target cell is in the clear but one group of edge cells are in the clutter cloud.

Sensitivity of Automatic-Threshold-Control Procedure to Changes in the First-Order Probability Distribution of Clutter Returns

A Rayleigh distribution has been assumed as a description of the first-order statistics of the clutter return in the threshold-control procedure (see Appendix IV). We now consider the case where a coherent component exists in, say, the ground clutter, and the probability density function is Rice distributed.

The false-alarm probability as a function of the ratio of coherent to incoherent back-scattering energy components of ground clutter is shown in Figure 16. In this sample computation, the threshold is originally set for a false-alarm probability of 10^{-8} when only incoherent scatterers exist. If a coherent component of the same average value for all the threshold control cells is introduced, a masking effect results, the threshold is raised, and the false-alarm probability is reduced.

A sample computation of the reverse situation follows. That is to say, the case where the threshold is initially set for 10^{-8} based on the assumption that an average coherent component of clutter energy exists in all the threshold control resolution cells and where the ratio of co-herent to incoherent component is, say, two. We then compute the value of the false-alarm probability for the case where the coherent component is not present and the distribution is Rayleigh. The false-alarm probability for this case rises to a value of 3.3×10^{-5}. It is clear that the false-alarm probability is sensitive to changes in the ratio of coherent to incoherent component of ground-clutter return. The false-alarm probability computations for these cases are based on the following argument.

We assume that the clutter return at the output of the ith resolution cell Z_i is a random variable with the Rayleigh distribution

$$f(Z_i)\,dZ_i = \frac{Z_i}{\sigma^2} \exp \left\{ \frac{-Z_i^2}{2\sigma_i^2} \right\} dZ_i \,.$$

With the addition of a coherent component, the output Z_i for this analysis is assumed to be Rice distributed.[15,16]

$$f(Z_i)\,dZ_i = \frac{Z_i}{\sigma^2} \exp \left\{ \frac{-Z_i^2 - S^2}{2\sigma^2} \right\} I_0 \left(\frac{SZ_i}{\sigma^2} \right) dZ_i \,, \qquad (49)$$

σ^2 is equal to the variance of each quadrature component and S is the coherent-component return. Introducing $Y_i = Z_i^2/(2\sigma^2)$,

$$f(Y_i)\,dY_i = \exp \left\{ - Y_i - X \right\} I_0 (2\sqrt{XY_i})\,dY_i \qquad (50)$$

where $X = S^2/(2\sigma^2)$ (the ratio of the energy in the coherent component to that in the incoherent component.)
The threshold

$$\beta_o = C \sum_{i=1}^{N} Y_i \,, \qquad (51)$$

where we assume a uniform target-scattering function so that the weights $\{W_i\}$ are unity. It is also assumed that the outputs (Y_i) are statistically independent and identically distributed in accordance with Equation (50).

The probability of false alarm in the cell under test is

$$P_{FA} = \int_{\beta_o=0}^{\infty} \int_{\beta_o=Y_o}^{\infty} \exp \left\{ - Y_o - X \right\} I_0 (2\sqrt{XY_o})\,dY_o f(\beta_o)\,d\beta_o \qquad (52)$$

where Y_o is the output of the cell under test, and the distribution of β_o is noted from Equation (51) to involve the convolution of the random

[15] D. E. Kerr, "Propagation of Short Wave Radios," *Radiation Lab. Ser.*, Vol. 13, McGraw-Hill, New York, 1951.
[16] P. Beckman, "Rayleigh Distribution and Its Generalizations," *Radio Sci.*, Nat. Bureau of Standards, Vol. 68D, Sept. 1964.

variables $\{Y_i\}$. This portion of the problem is the determination of the distribution of the sum of the square-law-detected outputs for a sine wave in noise and, from Marcum,[11] the distribution is

$$f(\beta_o)\,d\beta_o = \left(\frac{\beta_o}{CNX}\right)^{N-1} \exp\left\{-\frac{\beta_o}{C} - N\alpha\right\} I_{N-1}\left(2\sqrt{\frac{NX\beta_o}{C}}\right)\frac{d\beta_o}{C}.$$

(53)

An approximate solution to P_{FA} (Equation (52)) has been implemented on a computer and is the basis for the presented data (Figure 16). For the case where $X = 0$ (no coherent component), the false-alarm probability reverts to that of Equation (36) where

$$P_{FA} = \left[\cfrac{1}{1 + \cfrac{K^2}{N}}\right]^{N}, \qquad \text{and} \quad \frac{K^2}{N} = C.$$

The sensitivity of the automatic-threshold-control procedure to the factor $\alpha^2/2$, the ratio of coherent to incoherent return energy of the ground clutter, suggests the procedure of making the threshold proportional to a maximum-likelihood estimate of α. (It will be recalled that the developed procedure involves assuming a Rayleigh distribution of return, that is to say, the case where $\alpha = 0$.) We briefly explore the structure of such a procedure for the case where the value of α is the same for each resolution cell in the clutter region. The likelihood function is expressed as

$$L(Z_1, Z_2, Z_3 \ldots Z_n) = \prod_{i=1}^{N} Z_1 \exp\left\{-\frac{Z_i^2 + \alpha^2}{2}\right\}I_o(\alpha Z_i)\,dZ_i.$$

(54)

Setting $\partial \ln L/\partial \alpha = 0$,

$$0 = -\hat{a} + \sum_{i=1}^{N}\left[\frac{I_{-1}(\hat{a}Z_i) + (\hat{a}Z_i)}{2I_o(\hat{a}Z_i)}\right]Z_i,$$

(55)

which is an explicit relationship for the estimate $\hat{a}_1$. For the practical case where $\alpha Z_i > 1$, however, we can use the approximate relationship

$$\hat{a} \approx \sum_{i=1}^{N} Z_i.$$

(56)

This is a basically different result and suggests that the threshold be made proportional to the sum of linearly detected outputs of the surrounding resolution cells rather than the square-law-detected outputs. A further analysis of procedures of this type is required in order to accommodate the detection situation where coherent as well as incoherent clutter returns are present.
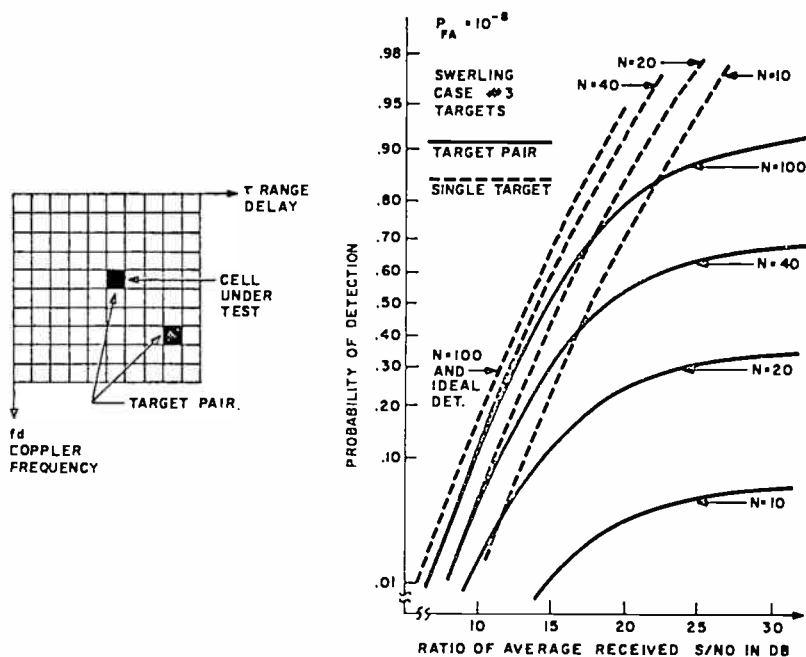
Fig. 18—Detection probability with automatic threshold control for a target pair model.

XI—PERFORMANCE WITH TARGET PAIR MODEL

An investigation of the performance of this procedure is now made for the case where instead of a single point target, the problem involves the detection of at least one of a pair of point targets not in the same resolution cell, but where both targets are embraced by the contiguously spaced resolution cells employed in the threshold-control function. It is intuitively clear that a masking effect results from this situation, for when the resolution cell containing one of the target pair is under test, the other target lodged in one of the threshold control cells provides a disproportionately high contribution to the threshold value. It is necessary, however, to quantitatively evaluate the loss of detectability due to this effect.

A set of computed results based on this analysis appear in Figure 18 for the case where both targets in the pair are fluctuating targets of the Swerling Case #3 first-order-amplitude distribution. The worst-case relationship occurs when the average target cross section is the same for the pair of targets. Since the detection mission involves alarming the system to the presence of at least one target at the beam position under test, detection would be made easier if the average cross section of one of the pair of targets predominated.

The family of computed curves in Figure 18 presents the detection probability as a function of the ratio of the average signal energy to noise-power density of each target of the pair. The number of resolution cells employed in the threshold control is an independent parameter, and the detection efficiency is an increasing function of this number. The intuitive explanation for this relationship is that as the number of resolution cell outputs contributing to the threshold value increases, the contribution of any single cell output to the final threshold value is reduced. It follows that the threshold control cell housing one of the target pair will make less of a masking effect contribution to the threshold setting. The masking effect may be introduced as a loss in detectability. As shown in Figure 18, with a detection probability of 0.5, a single-cell false-alarm probability of 10^{-8}, and 100 threshold control cells, the loss in detectability due to the target pair is approximately 1 dB.

XII—Detection Probability Formulation

The final case considered here is that of a point target embedded in an extended-clutter target cloud and where a second point target appears in one of the threshold control cells. The clutter target-scattering function will be assumed to be uniform over the breadth of the parameter space occupied by the threshold-control cells. The target-present hypothesis is accepted if the sampled square-law-detected output Y_o is greater than the threshold value T_o. The average signal-energy to clutter-noise-power density will be designated $\overline{X}_o$ for the target lodged in the cell under test and $\overline{X}_1$ for the target appearing in one of the threshold control cells. The probability of detection,

$$P_D = P\{E : (X_o > T_o) ; \overline{X}_o, \overline{X}_1\} , \tag{57}$$

is developed by initially considering the threshold value T_o, which is a random variable expressed for this case as

$$T_o = \frac{K^2}{N} \left[\sum_{i=1}^{N-1} Y_i + \theta_N \right] \tag{58}$$

The weights $\{W_i\}$ in Equation (1) are each equal to unity for the case at hand (a uniform distribution is assumed). The Y_i are the square-law-detected clutter return outputs of the $N-1$ threshold control cells not containing the point target, and the random variable θ_N represents the output of the threshold-control cell containing the second point target as well as the clutter return.

Performing a change of variable of both the outputs and the threshold with respect to the clutter power density, $\sigma_o^2 : \eta_i = Y_i/\sigma_o^2$, $\zeta_N = \theta_N/\sigma_o^2$ and $\beta' = T_o/\sigma_o^2$, the normalized threshold β' may be expressed as

$$\beta' = \frac{K^2}{N}\left[\sum_{i=1}^{N-1}\eta_i + \zeta_N\right].$$ (59)

The probability-density function of the normalized square-law-detected outputs $\{\zeta_i\}$ not including the target cell are exponentially distributed,

$$f(\eta_i)d\eta_i = \exp\{-\eta_i\}d\eta_i$$ (60)

for $i = 1, 2, \ldots N-1$. The probability-density function of the threshold-control cell containing the second target may be expressed

$$f(\zeta_N, \overline{X}_1)d\zeta_N = \int_{X=0}^{\infty} \exp\{-\zeta_N - X\}I_o(2\sqrt{X\zeta_N})f(X,\overline{X}_1)dXd\zeta_N \quad (61)$$

The expression

$$\exp\{-\zeta_N - X\}I_o(2\sqrt{X\zeta_N})\,d\zeta_N X$$ (62)

represents the conditional probability density function of the normalized square-law-detected output ζ_N on the hypothesis that the input signal-energy to clutter-power density is X. The probability-density function of X for the Swerling Case $\#$ 3 target is

$$f(X;\overline{X})\,dX = \frac{4X}{\overline{X}^2}e^{-2X/\overline{X}}dX$$ (63)

Performing the indicated integration of Equation (70),

$$f(\zeta_N, \overline{X}_1)\,d\zeta_N = \alpha_1^2 e^{-\alpha_1 \zeta_N}[1 + d_1\zeta_N]d\zeta_N.$$ (64)

where

$$\alpha_1 = \frac{1}{1 + \dfrac{\overline{X}_1}{2}} \quad \text{and} \quad d_1 = \frac{\overline{X}_1}{\overline{X}_1 + 2}$$

The development of the probability-density function of the normalized threshold β' is now required. A change of notation for Equation (68) is first introduced so that β' may be expressed as $\beta' = (Z + t)/\tau$ where

$$\tau = \frac{N}{K^2}, \quad Z = \sum_{i=1}^{N-1} \eta_i, \quad \text{and} \quad t = \zeta_N .$$

The probability density function of $Z = \sum_{i=1}^{N-1} \eta_i$ represents the convo-

lution of the exponentially distributed random variables $\{\eta_i\}$ and is shown to be

$$f_1(Z) dZ = \frac{Z^{N-2}}{(N-2)!} e^{-Z} dZ . \tag{65}$$

The density function of $l = Z + t$ may be expressed as the convolution

$$f(l) dl = \int_0^Z f_1(Z - t) f_Z(t) dt . \tag{66}$$

Introducing the change of variable $\beta' = Z/\tau$

$$\phi(\beta') = \int_0^{\tau\beta'} \tau\alpha_1^2 e^{-\alpha_1(\tau\beta'-t)} [1 + d_i(\tau\beta' - t)] \frac{t^{N-2}}{(N-2)!} e^{-t} dt \tag{67}$$

where

$$d_i = \frac{\overline{X}_i}{\overline{X}_i + 2}, \quad i = 0, 1 .$$

The detection probability P_D may now be expressed

$$P_D = \int_{\beta'=0}^{\infty} \alpha_o^2 e^{-\alpha_o \xi_o} [1 + d_o \zeta_o] d\zeta_o \tag{68}$$

where

$$P\{E : (\zeta_o > \beta')\} = \int_{e'}^{\infty} \alpha_o^2 e^{-\alpha_o \xi_o} [1 + d_o \zeta_o] d\zeta_o \tag{69}$$

and

$$P\{E : (\zeta_o > \beta')\} = e^{-\alpha_o \beta'} [1 + \alpha_o d_o \beta'] d\beta' \tag{70}$$

Introducing Equations (67) and (70) into the expression for the detection probability (Equation (57)), and performing the required integrations, the following result is obtained:

$$P_D = \frac{\tau^N \alpha_1^2}{(\tau\alpha_1 + \alpha_o)^2 (\tau + \alpha_o)^{N-1}} \left[\tau + \alpha_o + N\alpha_o d_2 + \frac{2\tau d_1 \alpha_o d_o}{\tau\alpha_1 + \alpha_o} \right], \tag{71}$$

where

$$\tau = \frac{N}{K^2}, \quad \alpha_i = \frac{2}{\overline{X}_i + 2}, \quad d_i = \frac{\overline{X}_i}{\overline{X}_i + 2} \qquad [i = 0, 1]$$

The performance results presented in Figure 18 are based on Equation (71). For the case where only one target is present, and it is in the cell under test, the expression for the detection probability of Equation (71) reduces to the form

$$P_D = \frac{\tau^N}{(\tau + \alpha_o)^{N+1}} [\tau + \alpha_o + N\alpha_o d_o] . \tag{72}$$

When expressed in terms of $\overrightarrow{X}$, this expression is noted to be the same as Equation (21).

For the case where the target pair is absent ($\overline{X}_1 = 0$ and $\overline{X}_2 = 0$),

the false-alarm probability (P_{FA}) is

$$P_{FA} = \left[\cfrac{1}{1 + \cfrac{K^2}{N}} \right]^N \tag{73}$$

This expression for the single-cell false-alarm probability is independent of the clutter level. For a specified single-cell false-alarm probability (P_{FA}), the threshold-control constant is computed from

$$K = [N((P_{FA})^{-1/N} - 1)]^{1/2} . \tag{74}$$

APPENDIX I—PROBABILITY OF DETECTION OF A NONFLUCTUATING TARGET WITH AUTOMATIC THRESHOLD CONTROL

The probability of detection of P_D is

$$P_D = \int_0^\infty f(\beta) Q(\alpha,\beta) \, d\beta , \tag{75}$$

where $Q(\alpha,\beta)$ is the conditional probability of detection on the hypothesis that the threshold is β, and the input signal energy to noise power density is $\alpha^2/2$ and

$$Q(\alpha,\beta) = 1 - \int_0^\beta X \exp\left\{ -\frac{X^2 + \alpha^2}{2} \right\} I_o(\alpha X) \, dX , \tag{76}$$

and where $f(\beta) d\beta$ is the threshold probability density function

$$f(\beta) \, d\beta = \frac{2}{\gamma(N-1)!} \left(\frac{\beta}{\gamma} \right)^{2N-1} \exp\left\{ -\left(\frac{\beta}{\gamma} \right)^2 \right\} \, d\beta \tag{77}$$

and where $\gamma^2 = 2K^2/N$.

The development begins with expressing the double integral of Equation (75) in the form

$$P_D = 1 - \int_0^\infty X \exp\left\{-\frac{\alpha^2 + X^2}{2}\right\} I_o(\alpha X) \left\{\int_X^\infty f(\beta)\,d\beta\right\} dX \quad (78)$$

$$\int_X^\infty f(\beta)\,d\beta = \frac{1}{(N-1)!} \int_X^\infty \left[\left(\frac{\beta}{\gamma}\right)^2\right]^{N-1} \exp\left\{-\left(\frac{\beta}{\gamma}\right)^2\right\}^2 d\left(\frac{\beta}{\gamma}\right) \quad (79)$$

$$\int_X^\infty f(\beta)\,d\beta = \frac{1}{(N-1)!} \int_{\frac{X}{\gamma}}^\infty (y^2)^{N-1} \exp\left\{-y^2\right\} dy^2 \quad (80)$$

$$\int_X^\infty f(\beta)\,d\beta = \frac{1}{(N-1)!} \int_{\left(\frac{X}{\gamma}\right)^2}^\infty Z^{N-1} e^{-Z}\,dZ \quad (81)$$

$$\int_X^\infty f(\beta)\,d\beta = \exp\left\{-\left(\frac{X}{\gamma}\right)^2\right\} \sum_0^{N-1}\left[\frac{1}{n!}\left(\frac{X}{\gamma}\right)\right] \quad (82)$$

$$P_D = 1 - \sum_0^{N-1} \frac{e^{-\alpha^2/2}}{\gamma^{2n} n!} \int_0^\infty X^{2n+1} \exp\left\{-X^2\left(\frac{1}{2} + \frac{1}{\gamma^2}\right)\right\} I_o(\alpha X)\,dX \ . \tag{83}$$

Let $\quad \tau^2 = \dfrac{1}{2} + \dfrac{1}{\gamma^2}$

$$P_D = 1 - \frac{1}{2\tau^2} \sum_n^{N-1} \frac{\exp\left\{-\dfrac{\alpha^2}{2}\right\}}{n!(\tau\gamma)^{2n}} \int_0^\infty (\tau X)^{2n} \exp\left\{-\gamma^2 X^2\right\} I_o\left(\frac{\alpha\tau X}{\tau}\right) 2\tau^2 X\,dX \ . \quad (84)$$

$$P_D = 1 - \frac{1}{2\tau^2} \sum_0^{N-1} \frac{\exp\left\{-\dfrac{\alpha^2}{2}\right\}}{n!(\tau\gamma)^{2n}} \int_0^\infty t^n e^{-t} I_o(2\sqrt{\xi t})\,dt \quad (85)$$

where $\quad \xi = \dfrac{\alpha^2}{4\tau^2} = \dfrac{\alpha^2\gamma^2}{2(\gamma^2 + 2)}$

$$P_D = 1 - \frac{\gamma^2}{\gamma^2 + 2} \sum_0^{N-1} \frac{\exp\{-\alpha^2/(\gamma^2 + 2)\}}{n!} \left(\frac{2}{\gamma^2 + 2}\right)^n e^{-\xi} \int_0^\infty t^n e^{-t} I_0(2\sqrt{\xi t})\, dt \; ; \quad (86)$$

leading to

$$P_D = 1 - \frac{\gamma^2}{\gamma^2 + 2} \sum_0^{N-1} \frac{\exp\{-\alpha^2/(\gamma^2 + 2)\}}{n!} \left(\frac{2}{\gamma^2 + 2}\right)^n L_n(\xi) \; . \quad (87)$$

where $L_m(\xi)$ are Laguerre polynominals defined by the relationships

$L_0(\xi) = 1, \quad L_1(\xi) = 1 + \xi, L_{m+1}(\xi) = (\xi + 2m + 1)L_m(\xi) - m^2 L_{m-1}(\xi)$

and

$$\xi = \frac{\alpha^2\gamma^2}{2(\gamma^2 + 2)}, \quad \gamma^2 = \frac{2k^2}{N}. \quad (88)$$

Appendix II—Probability of False Alarm in a Single Resolution Cell with Automatic Threshold Control

The single-cell probability of false alarm P_{FA} may be expressed as

$$P_{FA} = \int_0^\infty f(\beta) \int_\beta^\infty X \exp\left\{-\frac{X^2}{2}\right\} dX d\beta \quad (89)$$

where $f(\beta)\, d\beta$ is given by Equation (77).

The development begins with the performance of the integration with respect to X:

$$P_{FA} = \int_0^\infty \exp\left\{-\frac{\beta^2}{2}\right\} f(\beta)\, d\beta \quad (90)$$

$$P_{FA} = \frac{2}{(N-1)!} \int_0^\infty \left(\frac{\beta}{\gamma}\right)^{2N-1} \exp\left\{-\left(\frac{\beta}{\gamma}\right)^2\right\} \exp\left\{-\frac{\beta^2\gamma^2}{2\gamma^2}\right\} \frac{d\beta}{\gamma} \left\{\frac{\beta^2+\gamma^2}{2\gamma^2}\right\}$$

(91)

$$P_{FA} = \frac{2}{(N-1)!} \int_0^\infty t^{2N-1} \exp\left\{-t^2\left(1+\frac{\gamma^2}{2}\right)\right\} dt$$

(92)

$$P_{FA} = \frac{1}{(N-1)!} \left(\frac{2}{2+\gamma^2}\right)^N \int_0^\infty \left[\left[\frac{2+\gamma^2}{2}\right]t^2\right]^{N-1}$$

$$\exp\left\{-t^2\left(\frac{2+\gamma^2}{2}\right)\right\} 2\left(\frac{2+\gamma^2}{2}\right) t\,dt$$

(93)

$$P_{FA} = \frac{1}{(N-1)!} \left(\frac{2}{2+\gamma^2}\right)^N \int_0^\infty y^{N-1} \exp\left\{-y\right\} dy$$

(94)

$$P_{FA} = \left(\frac{2}{2+\gamma^2}\right)^N.$$

In terms of the original notation,

$$P_{FA} = \left[\frac{1}{1+\dfrac{K^2}{N}}\right]^N.$$

(95)

Appendix III—Probability of Detection of a Swerling Case #3
Target with Automatic Threshold Control

The detection probability (P_D) may be expressed

$$P_D = \int_{\beta=0}^\infty f_1(\beta) \int_{\alpha=0}^\infty Q(\alpha,\beta) f_2(\alpha;\bar{X})\,d\alpha d\beta,$$

(100)

where $f_1(\beta)d\beta$ is defined by Equation (77) and for the Swerling Case #3 target,

$$f_2(\alpha;\overline{X})d\alpha = \frac{2\alpha^3}{\overline{X^2}}\exp\left\{-\frac{\alpha^2}{\overline{X}}\right\}d\alpha.$$

Performing the integration with respect to α first:

$$P_D = \int_{\beta=0}^{\infty} f_1(\beta)\exp\left\{-\frac{\beta^2}{\overline{X}+2}\right\}\left[1+\frac{\beta^2\overline{X}}{(\overline{X}+2)^2}\right]d\beta \quad (101)$$

$$P_D = \int_{0}^{\infty}\frac{2}{(N-1)!}\left(\frac{\beta}{\gamma}\right)^{2N-1}\exp\left\{-\left(\frac{\beta}{\gamma}\right)^2\right\}$$

$$\exp\left\{-\frac{\gamma^2}{\overline{X}+2}\left(\frac{\beta}{\gamma}\right)^2\right\}\left[1+\frac{\left(\frac{\beta}{\gamma}\right)^2\gamma^2\overline{X}}{(\overline{X}+2)^2}\right]d\left(\frac{\beta}{\gamma}\right). \quad (102)$$

Let $(\beta/\gamma)^2 = X$,

$$P_D = \frac{1}{(N-1)!}\int_{0}^{\infty}X^{N-1}\exp\{-X\}\exp\left\{-\frac{\gamma^2}{\overline{X}+2}X\right\}\left[1+\frac{\gamma^2\overline{X}}{(\overline{X}+2)^2}X\right]dX$$

$$\text{(103)}$$

$$P_D = \frac{1}{(N-1)!}\left(\frac{\overline{X}+2}{\overline{X}+2+\gamma^2}\right)^N\int_{0}^{\infty}y^{N-1}\exp\{-y\}\left[1+\frac{\gamma^2\overline{X}}{(\overline{X}+2)(\overline{X}+2+\gamma^2)}\right]dy$$

$$\text{(104)}$$

$$P_D = \left(\frac{\overline{X}+2}{\overline{X}+2+\gamma^2}\right)^N\left[1+\frac{N\gamma^2\overline{X}}{(\overline{X}+2)(\overline{X}+2+\gamma^2)}\right]. \quad (105)$$

Substituting $\gamma^2 = \dfrac{2K^2}{N}$,

$$P_D = \left[\frac{\overline{X} + 2}{\overline{X} + 2 + \dfrac{2K^2}{N}} \right]^N \left[1 + \frac{2K^2\overline{X}}{(\overline{X} + 2)\ \left(\overline{X} + 2 + \dfrac{2K^2}{N}\right)} \right] \quad (106)$$

It is interesting to note that as $N \rightarrow \infty$,

$$P_D = \exp \left\{ -\frac{2K^2}{\overline{X} + 2} \right\} \left[1 + \frac{2K^2\overline{X}}{(\overline{X} + 2)^2} \right] ,$$

which is the P_D for the Swerling Case #3 target with a fixed threshold and known noise power density when $\beta^2 = 2K^2$.

Appendix IV—Detection Probability for the Swerling Case #3 Target and False-Alarm Probability Where a Target-Scattering Function Shape Mismatch Exists

The probability of detecting a Swerling Case #3 target is developed in this Appendix for the case where the backscattering clutter target scattering function assumed in designing the weights employed in the automatic threshold procedure differs from the clutter target-scattering function actually encountered.

This probability is

$$P_D = \frac{N}{K^2} \sum_{J=1}^{N} \frac{\eta_J}{\gamma_J} \left[1 + \frac{\overline{X}}{(\overline{X} + 2)^2\gamma_J} \right] \quad (107)$$

where

$$\gamma_J = \frac{2}{\overline{X} + 2} + \frac{\alpha_J N}{K^2}$$

$$\eta_J = \frac{\displaystyle\prod_{i=1}^{N} \alpha_i}{\displaystyle\prod_{i \neq J} (\alpha_i - \alpha_J)} \quad (108)$$

and $N =$ total number of resolution cells surrounding the cell under test and employed in threshold control.

$\overline{X} =$ average signal energy to total noise power density at the cell under test.

$\alpha_J =$ the ratio at the Jth cell of the assumed noise power density to the 'true' value of the noise power density $(\alpha_J = \sigma_J{}^2/\underline{\sigma_J{}^2})$.

$K =$ the threshold control constant (see Equation (3)).

We first develop the probability density function for the normalized threshold β (i.e., $f(\beta)d\beta$). β is defined as

$$\beta = \frac{K}{\sigma_o} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \frac{X_i{}^2}{W_i}} \tag{109}$$

where $W_i = \sigma_i{}^2/\sigma_o{}^2$, the ratio of the assumed variance at the ith resolution cell to that at the target cell.

The probability density function of the output of the ith resolution cell X_i is assumed to be Rayleigh distributed:

$$f(X_i)\,dX_i = \frac{X_i}{\underline{\sigma_i{}^2}} \exp\left\{-\frac{X_i{}^2}{2\underline{\sigma_i{}^2}}\right\} dX_i. \tag{110}$$

and where the sub-bar refers to the 'true' value $\underline{\sigma_i{}^2}$. Introducing the change of variable $Z_i = X_i{}^2/(2\sigma_i{}^2)$, Equations (109) and (110) become, respectively,

$$\beta = K \sqrt{\frac{2}{N}} \sqrt{\sum_{i=1}^{N} Z_i} \tag{111}$$

and

$$f(Z_i)\,dZ_i = \alpha_i \exp\{-\alpha_i Z_i\}\,dZ_i \tag{112}$$

where

$$\alpha_i = \frac{\sigma_i{}^2}{\underline{\sigma_i{}^2}} \tag{113}$$

We first develop the distribution of the random variable

$$l = \sum_{i=1}^{N} Z_i .$$

Taking the Laplace transformation of $f(l)$ and making a partial fraction expansion we have

$$L(f(l)) = \prod_{i=1}^{N} \alpha_i \prod_{i=1}^{N} \frac{1}{S + \alpha_i} = \sum_{J=1}^{N} \frac{\eta_J}{S + \alpha_J} \tag{114}$$

where η_J is defined as in Equation (108). Taking the inverse transformation,

$$f(l) \, dl = \sum_{J=1}^{N} \frac{\prod\limits_{i=1}^{N} \alpha_i \exp \{- \alpha_J l\}}{\prod\limits_{i \neq J} (\alpha_i - \alpha_J)} \, dl \tag{115}$$

We now wish the distribution of $T = \sqrt{l}$. Introducing this change of variable in Equation (115),

$$f(T) dT = 2 \sum_{J=1}^{N} T \eta_J \exp \{- \alpha_J T^2\} \, dT \tag{116}$$

Since $\beta = K\sqrt{2/N} \, T$, we now introduce this relationship as a change of variable in Equation (11) and obtain

$$f(\beta) \, d\beta = \frac{N}{K^2} \sum_{J=1}^{N} \eta_J \beta \exp \left\{ - \frac{\alpha_J N \beta^2}{2K^2} \right\} d\beta . \tag{117}$$

Using the developed probability density function of the normalized threshold (Equation (117)), we are now in a position to develop the probability of detection for the Swerling Case #3 target as in Appendix III.

Performing the integration first with respect to α,

$$P_D = \int_{\beta=0}^{\infty} f_1(\beta) \left[\exp \left\{ - \frac{\beta^2}{\overline{X} + 2} \right\} \left[1 + \frac{\beta^2 \overline{X}}{(X + 2)^2} \right] \right] d\beta , \tag{118}$$

leading to the final result,

$$P_D = \frac{N}{K^2} \sum_{J=1}^{N} \frac{\eta_J}{\gamma_J} \left[1 + \frac{\overline{X}}{(\overline{X}+2)^2 \gamma_J} \right] \quad (119)$$

where γ_J and η_J are defined by Equations (108)

The false alarm probability is developed from Equation (119) simply by setting $\overline{X} = 0$, yielding

$$P_{FA} = \frac{N}{K^2} \sum_{J=1}^{N} \frac{\eta_J}{\left[1 + \frac{N\alpha_J}{K^2} \right]}. \quad (120)$$
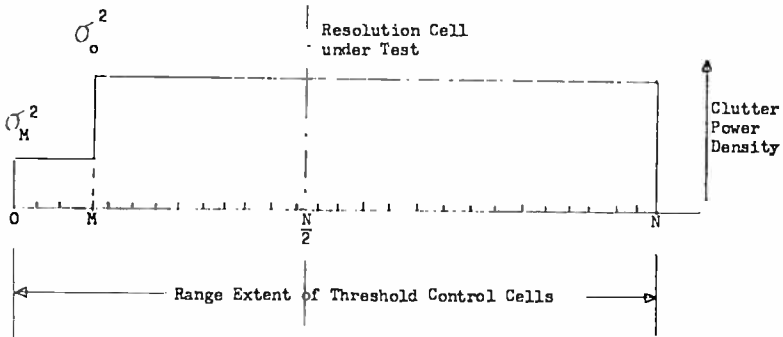
Fig. 19

APPENDIX V—SWERLING CASE #3 DETECTION AND FALSE ALARM PROBABILITY WHERE A STEP-FUNCTION DISCONTINUITY OF CLUTTER DISTRIBUTION IN RANGE (EDGE AFFECT ANALYSIS)

In this appendix, the probability of detecting a Case #3 target type is developed for the case where the target-scattering function is assumed to be distributed uniformly in range but where the actual distribution can be represented as having a step function discontinuity of the type illustrated in Figure 19.

The developed expression for the detection probability and the false-alarm probability (P_{FA}) are determined to be,

$$P_D = \left[\frac{\gamma}{\alpha + \gamma} \right]^M \frac{\left(1 + \beta \left(\frac{M}{\alpha + \gamma} + \frac{N - M}{1 + \alpha} \right) \right)}{(1 + \alpha)^{N - M}} \tag{121}$$

where

$$\beta = \frac{2K^2 \overline{X}}{N(\overline{X} + 2)^2}$$

$$\alpha = \frac{2K^2}{N(\overline{X} + 2)} \; ;$$

and

$$P_{FA} = \frac{\left[\dfrac{\gamma}{\dfrac{K^2}{N} + \gamma} \right]^M}{\left(1 + \dfrac{K^2}{N} \right)^{N - M}} \tag{122}$$

where $\gamma = \sigma_b^2 / \sigma_a^2$. γ is the ratio of the variance of the total noise at the edge to that at the cell under test. Performing the integration of the detection probability first with respect to α, yields

$$P_D = \int_{\beta=0}^{\infty} f_1(\beta) \left[\exp \left\{ -\frac{\beta^2}{\overline{X} + 2} \right\} \left[1 + \frac{\beta^2 + \overline{X}}{(\overline{X} + 2)^2} \right] \right] d\beta \tag{123}$$

The normalized threshold β may be expressed in general as

$$\beta = \frac{K}{\sigma} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \frac{X_i^2}{W_i^2}} \tag{124}$$

and for the case at hand, a uniform distribution in range is assumed so that the weights W_i are unity.

The linearly detected outputs of the matched filters surrounding the resolution cell under test are Rayleigh distributed:

$$f(X_i)\,dX_i = \frac{X_i}{\sigma_i^2}\exp\left\{-\frac{X_i^2}{2\sigma_i^2}\right\}dX_i\,, \tag{125}$$

where $\sigma_i^2 = \sigma_E^2$ for $i \leqslant M$ and $\sigma_i^2 = \sigma_o^2$ for $i > M$. For both cases $M \leqslant (N/2) - 1$. If we introduce the change of variable $Z_i^2 = X_i^2/(2\sigma_i^2)$ in Equation (125), we have:

$$f(Z_i)\,dZ_i = \gamma_i\,e^{-\gamma_i Z_i}dZ_i \tag{126}$$

where

$$\gamma_i = \frac{\sigma_o^2}{\sigma_E^2} = \gamma_o \ \text{ for } \ i \leqslant M$$

$$\gamma_i = \frac{\sigma_o^2}{\sigma_o^2} = 1 \ \text{ for } \ i > M$$

and where for both cases $M < (N/2) - 1$.

We now are ready to introduce the change of variable $\beta^2 = (2K^2/N)\,t$ into Equation (7) where

$$t = \sum_{i=1}^{N} Z_i \tag{127}$$

The expression for the detection probability then takes the form

$$P_D = \int_{t=0}^{\infty} f(t)\,\exp\left\{-\frac{2K^2t}{N(\bar{X}+2)}\right\}\left[1 + \frac{2K^2\vec{X}t}{N(\vec{X}+2)^2}\right]dt \tag{128}$$

or

$$P_D = \int_{t=0}^{\infty} f(t)\,\exp\left\{-\alpha t\right\}[1 + \beta t]dt\,. \tag{129}$$

Our problem now involves finding the form of the density function of $t = \sum_{i=1}^{N} Z_i$. The Laplace transform of $f(t)$ is

$$L(f(t)) = \frac{\gamma_o{}^M}{(S + \gamma_o)^M (S + 1)^{N-M}} \tag{130}$$

If we define $F_1(S) = \dfrac{\gamma_o{}^M}{(S + \gamma_o)^M (S + 1)^{N-M}}$ \tag{131}

$$F_2(S) = \frac{1}{(S + 1)^{N-M}} \tag{132}$$

then

$$f(t)\,dt = \int_0^t f_1(t - \tau) f_2(\tau)\,d\tau \tag{133}$$

where

$$f_1(t) = L^{-1}[F_1(S)] = \frac{\gamma^M t^{M-1}}{(M - 1)!} e^{-\gamma t} \tag{134}$$

and

$$f_2(t) = L^{-1}[F_2(S)] = \frac{t^{N-M-1}}{(N - M - 1)!} e^{-t}. \tag{135}$$

Now substituting Equation (133) into Equation (129),

$$P_D = \int_0^\infty e^{-\alpha t}(1 + \beta t) \int_0^t f_1(t - \tau) f_2(\tau)\,d\tau dt \tag{136}$$

$$= \int_\alpha^\infty \int_\tau^\infty e^{-\alpha(t-\tau)}\ e^{-\alpha\tau}[1 + \beta\tau + \tau(t - \tau)]f_1(t - \tau) f_2(\tau)\,d(t - \tau)d\tau \tag{137}$$

$$P_D = \int_0^\infty e^{-\alpha\tau} f_2(\tau) \int_0^\infty e^{-\alpha t}(1 + \beta\tau + \beta t)f_1(t)\,dt\,d\tau \qquad (138)$$

$$= \int_0^\infty e^{-\alpha\tau} f_2(\tau)\, R(\tau)\, d\tau \qquad (139)$$

where

$$R(\tau) = \frac{\gamma^M}{(M-1)!} \int_0^\infty t^{M-1}(1 + \beta\tau + \beta t)\, e^{-(\alpha+\gamma)t}\, dt \qquad (140)$$

$$R(\tau) = \left(\frac{\gamma}{\alpha+\gamma}\right)^M \left(1 + \frac{\beta M}{\alpha+\gamma} + \beta\tau\right) \qquad (141)$$

$$P_D = \left(\frac{\gamma}{\alpha+\gamma}\right)^M \frac{1}{(N-M-1)!} \int_0^\infty t^{N-M-1}e^{-(1+\alpha)t}\left(1 + \frac{\beta M}{\alpha+\gamma} + \beta t\right) dt \qquad (142)$$

$$P_D = \left(\frac{\gamma}{\alpha+\gamma}\right)^M \frac{1}{(1+\alpha)^{N-M}} \int_0^\infty \frac{t^{N-M-1}\left(1 + \frac{M}{\alpha+\gamma} + \frac{\beta t}{1+\alpha}\right)e^{-t}dt}{(N-M-1)!} \qquad (143)$$

and finally

$$P_D = \left[\frac{\gamma}{\alpha+\gamma}\right]^M \frac{1 + \beta\left(\dfrac{M}{\alpha+\gamma} + \dfrac{N-M}{1+\alpha}\right)}{(1+\alpha)^{N-M}} \qquad (144)$$

where

$$\beta = \frac{2K^2\overline{X}}{N(\overline{X}+2)^2}, \quad \alpha = \frac{2K^2}{N(\overline{X}+B)}, \quad \gamma = \frac{\sigma_E{}^2}{\sigma_o{}^2}.$$

RCA Technical Papers

Second Quarter, 1968

"Ferroelectric Properties of Stable and Metastable Phase-III KNO_3," G. W. Taylor and B. J. Lechner, *Jour. Appl. Phys.* (April) .. 1968

"GaAs Laser Amplifiers," W. F. Kosonocky and R. H. Cornely, *IEEE Jour. Quantum Electronics,* (April) 1968

"Inexpensive Rubidium-D_2 Filter," A. H. Firester, *Amer. Jour. Phys.* (Notes and Discussion) (April) 1968

1967 Semiconductor Laser Conference "Introduction," J. I. Pankove, *IEEE Jour. Quantum Electronics* (April) 1968

Author's Reply to Comment on "An Introduction to Computer Graphic Terminals," M. H. Lewin, *Proc. IEEE* (Letters) (April) 1968

"Multilayer GaAs Injection Laser," W. F. Kosonocky, R. H. Cornely, and I. J. Hegyi, *IEEE Jour. Quantum Electronics* (April) .. 1968

"RF Furnace for Maintaining Temperature Differences Between Parallel Susceptors," L. A. Murray, P. A. Hoss, and J. J. Rivera, *Rev. Sci. Instr.* (Notes) (April) 1968

"Saturation of Irradiance Fluctuations Due to Turbulent Atmosphere," D. A. DeWolf, *Jour. Opt. Soc. Amer.* (April) 1968

"Stimulated Emission from Bulk Field-Ionized GaAs," P. D. Southgate, *IEEE Jour. Quantum Electronics* (April) 1968

"Temperature Dependence of Emission Efficiency and Lasing Threshold in Laser Diodes," J. I. Pankove, *IEEE Jour. Quantum Electronics* (April) 1968

"Time-Delay and Memory Effects in $GaAs_{1-x}P_x$ Injection Laser," J. I. Pankove, *IEEE Jour. Quantum Electronics* (April) 1968

"The Truth About Stereo Amplifier Power Ratings," R. S. Hartz and F. S. Kamp, *Audio* (April) 1968

"Field Effect on the Cochran Modes in $SrTiO_3$ and $KTaO_3$," E. F. Steigmeier, *Phys. Rev.* (10 April) 1968

"Superconductivity in Granular Aluminum Films," R. W. Cohen and B. Abeles, *Phys. Rev.* (10 April) 1968

"Airborne Military Transceiver Finds Room in Crowded Spectrum," I. P. Magasiny, *Electronics* (April 15) 1968

"Amplifier Recovery from Saturation Slowing You Down? Give It the Gate," H. P. Brockman, *Electronic Engineer* (May) 1968

"Carrier Transport in Thin Silicon Films," D. J. Dumin and P. H. Robinson, *Jour. Appl. Phys.* (May) 1968

"Cryoelectric Content-Addressable Memories Using Flux-Shuttling Nondestructive Readout," S. A. Keneman, *Jour. Appl. Phys.* (May) 1968

"Device Applications of Superconductive Tunneling," B. N. Taylor, *Jour. Appl. Phys.* (May) 1968

"High-Speed Reactive Q-Switching in CO_2 Lasers," A. I. Carswell and J. I. Wood, *IEEE Jour. Quantum Electronics* (Correspondence) (May) 1968

"The Pinning Potential and High-Frequency Studies of Type-II Superconductors," J. I. Gittleman and Coauthor, *Jour. Appl. Phys.* (May) 1968

"A Simple Display for Characters and Graphics," J. C. Miller and C. M. Wine, *Trans. IEEE PTGC* (May) 1968

"Studies of the Focal Region of a Spherical Reflector: Geometric Optics," G. Hyde and Coauthor, *Trans. IEEE PTGAP* (May) 1968

AUTHORS

HOWARD R. BEELITZ received the B.S.E.E. degree from Newark College of Engineering in 1959 and has since done graduate work at Newark College of Engineering and at Princeton University. Mr. Beelitz initially joined RCA Electronic Components, Harrison, N.J., in 1957 to work on the development of ceramic-to-metal seals for use in pencil-type vacuum tubes. In June, 1959, he joined the technical staff at RCA Laboratories. His early work involved both applications and fabrication studies of tunnel diodes. He was later engaged in research on thin-film organic diode and thick film resistor array associative memories. More recent assignments included research in monolithic sense circuitry for integrated ferrite memories. His current interests include large-scale integration of bipolar logic for experimental computers and computer aided interactive graphic techniques for artwork generation. Mr. Beelitz is a member of Tau Beta Pi and Eta Kappa Nu.

WOLFRAM A. BÖSENBERG received the M.S. and Ph.D. degrees in physics from the Technical University in Stuttgart, Germany in 1951 and 1955, respectively. From 1952 to 1957 he worked at Standard Elektrik Lorenz A.G., Nuremberg, Germany, on semiconductor material and device technology. In 1957 Dr. Bösenberg joined the RCA Electronic Components division at Somerville, New Jersey. He has participated in the development of various silicon and gallium arsenide devices. From 1960 to 1963 he held the positions of Manager, Advanced Technology, and Manager, Advanced Devices, and was responsible for the applied research and development of a wide variety of advanced semiconductor devices. In May 1963, Dr. Bösenberg transferred to RCA Laboratories, where he has worked on a number of field-effect integrated devices for computer applications. In 1966 he became head of the Integrated Circuits Facility and is presently responsible for the development and fabrication of a wide variety of advanced bipolar and field-effect integrated circuits.

467

PAUL BURA received the BSEE and MSEE degrees from the University of London in 1948 and 1949, respectively, and was awarded Diploma of Imperial College of Science and Technology. He is currently working toward a Ph.D. degree. From 1949 to 1954 he was a research engineer at Research Laboratories, A.E.I., England, where he worked on the development of magnetrons, TR and ATR tubes, and photomultipliers. He joined the Canadian General Electric Co. in 1954 and for the next three years he worked on airborne radar and multichannel communications equipment. At Canadian Westinghouse during the period 1957-1960 he developed troposcatter equipment for the 4400-5000 MHz range, and the parametric amplifier for the AN/TRC-66 equipment. From 1960 to 1962 he was a technical scientist at ITT Laboratories in charge of advanced development programs on parametric amplifiers. Mr. Bura joined RCA Defense Communications Systems Division in 1962. Since joining RCA he has worked on solid-state research, superconductivity, and parametric amplifiers. His latest work is in the area of microwave integrated circuits.

HAROLD M. FINN received the BEE degree from the City College of N. Y. 1950, the M.S. degree in E.E. from the University of Maryland in 1955, and the Ph.D. degree from the University of Pennsylvania in 1965. Since 1950, Dr. Finn has been engaged in research and development of industrial control, radar, and communications systems while employed at the Engineering and Research Corp., Melpar, and the Emerson Research Laboratories. He joined the Missile and Surface Radar Division of RCA in 1959 where he is engaged in the development and application of signal detection theory. He is presently with the Technology Service Corp., Santa Monica, California. Dr. Finn is a member of Tau Beta Pi, Eta Kappa Nu, and is a senior member of the

DORIS FLATLEY received the B.S. degree in chemical engineering from Newark College of Engineering in 1961. In 1968 she received an M.S. degree in engineering science from the same institution. From 1960 to 1964 she was employed by RCA Electronic Components, Somerville, engaged in research and development of gallium arsenide junction devices. In 1964 she transferred to RCA Laboratories, Princeton, N.J., and has been involved since that time in the development of field-effect integrated circuits. She is a member of the Electrochemical Society.

JOHN T. GRABOWSKI received the B.S.E.E. degree from the Massachusetts Institute of Technology in 1963 and M.S.E.E. and M.S. in Physics degrees from the University of Pennsylvania in 1966 and 1968, respectively. His program of study emphasized the areas of digital logic design, semiconductor physics, devices, and applications. He joined the technical staff of RCA Laboratories in 1963 where he has been engaged in research in the fields of semiconductor devices, nondestructive read-out memory development, and high-speed memory development. Mr. Grabowski is a Member of IEEE, Sigma Xi, Tau Beta Pi, and Eta Kappa Nu.

ROBERT S. JOHNSON received his B.S. degree in mathematics in 1950 and his M.S. degree in mathematics in 1951, both from Northwestern University. In 1953, he joined the Institute for Cooperative Research at the University of Pennsylvania where he was assigned to projects BIGBEN and CARAMU. He received his Ph.D. degree in mathematics from the University of Pennsylvania in 1959. He joined RCA in 1959 as a numerical analyst and scientific programmer, and has worked primarily in the areas of orbital mechanics, optimization techniques, and systems simulation. He was responsible for the numerical analysis and programming of the PAGE, LEMIP, and EAST programs. He is a member of the American Mathematical Society and of S.I.A.M.

FREDERICK H. NICOLL received the B.Sc. degree in physics from Saskatchewan University, Canada, in 1929 and the M.Sc. degree in 1931. He held an 1851 Exhibition Scholarship to Cambridge University, England, for three years research and received the Ph.D. degree from that university in 1934. He was a research physicist with Electric and Musical Industries. Ltd., in London from 1934 to 1939. From 1939 to 1941, he was with the Consumer Electronics Division at Camden, N.J., as a research engineer. Since 1942, he has been with RCA Laboratories in Princeton, N.J. engaged in research on cathode-ray tubes and electron optics and, more recently, photoconduction, electroluminescence, and solid-state devices. In 1963-64, Dr. Nicoll spent a year of study at the Cavendish Laboratory, Cambridge, England.

Dr. Nicoll is a Fellow of the Institute of Electrical and Electronics Engineers and a Member of the American Physical Society and of Sigma Ki.

ARNOLD SCHMIDT received the M.E.E. degree from the City College of New York in 1962. He joined the RCA Defense Communications Systems Division in the same year. He has worked on various aspects of data transmission in wire and radio systems, including an investigation of the pseudorandom and time-frequency coded RADAS approaches to multiple-access satellite communications, studies of PSK and FSK modulation techniques for HF links, analyses of signal detection in noise and probabilistic self-test methods, and investigation and analysis of new diversity techniques to combat correlated fading in troposcatter communications. In 1968, he joined the Communications Systems Programs Department of TRACOR, Inc., in New York. Mr. Schmidt is a Member of the Institute of Electrical and Electronics Engineers.

J. T. WALLMARK (See RCA REVIEW, Vol. 29. No. 2, June 1968.)