RCA REVIEW

a technical journal

RADIO AND ELECTRONICS
RESEARCH • ENGINEERING

RCA REVIEW

a technical journal

RADIO AND ELECTRONICS

RESEARCH • ENGINEERING

Published quarterly by

RCA LABORATORIES

in cooperation with all subsidiaries and divisions of

RADIO CORPORATION OF AMERICA

CONTENTS

RCA REVIEW

BOARD OF EDITORS

REPUBLICATION AND TRANSLATION

NONLINEAR-CAPACITANCE AMPLIFIERS*

BY

L. S. NERGAARD

RCA Laboratories,
Princeton, N. J.

Summary—This paper explains the operations of variable-capacitance amplifiers in simple physical terms. It then describes the forms which these amplifiers have taken to date. Finally, published operating data on variable-reactance amplifiers are reviewed and the effective noise temperatures achieved are compared with the noise temperatures of other low-noise amplifiers and with the noise temperatures of terrestrial and extra-terrestrial noise sources.

INTRODUCTION

RECENTLY there has been tremendous interest in a class of amplifiers commonly called parametric amplifiers. Interest in these amplifiers stems from theoretical predictions that they should exhibit very low noise and from the present urgency for the achievement of amplifiers with very low noise figures. Parametric amplifiers, properly, variable-reactance amplifiers, achieve gain by periodically varying a reactance in a resonant signal circuit. Such amplifiers are by no means new. In 1936, Hartley published a paper which deals with what would now be called parametric amplifiers.[1] Other early contributions are those of Manley and Peterson,[2] Torrey and Whitmer,[3] van der Ziel,[4] and Landon.[5] Of the recent theoretical studies those of Manley and Rowe,[6] Bloom and Chang,[7] and Heffner and Wade[8] are of immediate pertinence to a discussion of nonlinear-capacitance amplifiers. Experimental work will be considered in due course.

Three of the above references deserve special mention because they describe the how, why, and how much of nonlinear-capacitance ampli-

* Manuscript received December 22, 1958.

[1] R. V. L. Hartley, "Oscillations in Systems with Non-Linear Reactances," *Bell Sys. Tech. Jour.*, Vol. 15, p. 424, July, 1936.

[2] J. M. Manley and E. Peterson, "Negative Resistance Effects in Saturable Reactor Circuits," *Trans. A.I.E.E.*, Vol. 65, p. 870, December, 1946.

[3] H. C. Torrey and C. A. Whitmer, *Crystal Rectifiers*, MIT Rad. Lab. Series, Vol. 15, McGraw-Hill, New York, 1948.

[4] A. van der Ziel, "On the Mixing Properties of Non-Linear Capacitances," *Jour. Appl. Phys.*, Vol. 19, p. 999, November, 1948.

fiers. Torrey and Whitmer observed negative conversion conductances in crystal diode mixers and, with an assist from W. C. Hahn, showed that these negative conductances arise from the voltage dependence of the junction capacitance of a crystal diode. They worked out the theory of reactance-variation amplification in a form remarkably similar to recent theory and applied the theory to explain their experimental results in detail.

Van der Ziel concluded that a variable-reactance amplifier should exhibit very low noise. In brief, the argument is that no tubes with thermionic emitters are involved so there is no shot-effect noise. Ideally, the circuit contains only reactances so there is no Nyquist–Johnson noise. If there is any noise, it is due solely to the Nyquist–Johnson noise in the unavoidable resistances in the circuits.

Manley and Rowe established the energy relations in networks containing nonlinear reactances in very general terms. For example, they determined the fraction of the power input at one frequency which can be converted to output power at another frequency.

To sum up, van der Ziel put his finger squarely on the reason for present interest in variable-reactance amplifiers, i.e., their low noise; Torrey and Whitmer demonstrated the means by which such amplifiers can be achieved at microwave frequencies, i.e., the crystal diode; and Manley and Rowe established what the efficiency of such amplifiers can be.

For those who take delight in such things, it is noted that the differential equations which describe parametric amplifiers were investigated by Mathieu in 1868,[9] and Lord Rayleigh studied mechanical parametric amplifiers in 1883.[10]

The remainder of this paper discusses, in turn, the gain mechanism in nonlinear-capacitance amplifiers, diode capacitors, simple parametric amplifiers, the performance of present amplifiers and, finally, their status relative to other low-noise amplifiers.

---

[5] V. D. Landon, "The Use of Ferrite-Cored Coils as Converters, Amplifiers, and Oscillators," *RCA Review*, Vol. X, p. 387, September, 1949.

[6] J. M. Manley and H. E. Rowe, "Some General Properties of Non-Linear Elements," *Proc. I.R.E.*, Vol. 44, p. 904, July, 1956.

[7] S. Bloom and K. K. N. Chang, "Theory of Parametric Amplification Using Nonlinear Reactances," *RCA Review*, Vol. 18, p. 578, December, 1957.

[8] H. Haffner and G. Wade, "Gain, Band-Width, and Noise Characteristics of a Variable Parameter Amplifier," *Jour. Appl. Phys.*, Vol. 29, p. 1321, September, 1958.

[9] E. T. Whittaker and G. N. Watson, *Modern Analysis*, Cambridge University Press, Cambridge, England, 1927.

[10] Lord Rayleigh, "On Maintained Vibrations," *Phil. Mag.*, Vol. 15, p. 229, 1883.

Gain Mechanism in Variable-Reactance Amplifiers

Before delving into the variety of nonlinear-capacitance amplifiers now under study, it is desirable to have a clear physical picture of the gain mechanism by which they operate. To establish this physical picture, consider the simple circuit of Figure 1. The circuit consists of a tank circuit with a capacitor whose plates can be separated mechanically. Assume that prior to time $t = 0$, the circuit has been excited so that the voltage V across the capacitor and the charge Q on the capacitor are varying sinusoidally. The curves in the lower part of Figure 1 show the charge and voltage as they are at time $t = 0$ and

Fig. 1—Charge and voltage build-up in a resonant circuit when the capacitance is varied at twice the resonant frequency.

during the quarter period following. To obtain amplification, the capacitor plates are pulled apart when the charge and voltage are at their maxima. Because of the electric field between the capacitor plates, it requires an expenditure of mechanical energy to pull the capacitor plates apart. This mechanical energy appears as additional electrical energy stored in the capacitor and manifests itself as an abrupt increase in the voltage as shown in the lower part of the figure. The voltage and charge then continue their oscillation towards zero and at zero voltage the capacitor plates are restored to their initial spacing. Restoring the plates at this time requires no work because there is no field between plates, hence no force. The voltage and charge then swing to

their negative maxima where the capacitor plates are again abruptly separated, and so on. Each time the capacitor plates are separated, energy is added to the signal so that the amplitude builds up. If the plates were separated by the same amount each half cycle, the amplitude would build up indefinitely. However, as the amplitude builds up it requires more and more force to separate the plates, so ultimately the force required would be infinite. With only a finite force available, the amplitude builds up asymptotically to a finite value where the energy added per separation just equals the energy dissipated per half cycle.

Fig. 2—Waveforms showing the role of the idling circuit in a parametric amplifier.

It will be noted that in the present case, energy is added every half cycle, i.e., the circuit is "pumped" at twice the signal frequency, and that the phase of the pump is important. There are many cases where it is desirable to pump at a frequency other than twice the signal frequency or to take the output at a frequency other than the signal frequency as in a converter, for example. This may be done if a second tuned circuit is coupled to the signal circuit by the variable capacitor as shown in the circuit of Figure 2. Here the signal circuit is tuned to the signal frequency and the second circuit, usually called the idling

circuit, is tuned to a frequency equal to the difference between the pumping frequency and the signal frequency. The signal circuit has a circulating current i_s which contributes a charge Q_s to the variable capacitor. The idling circuit has a circulating current i_i which contributes a charge Q_i. These charges are shown in the upper two curves of Figure 2. The total charge across the capacitor is $Q_s + Q_i$ as shown by the third curve in Figure 2. The fourth curve shows the variation of the coupling capacitance at the pump frequency. An examination of third and fourth curves of Figure 2 shows that if the variable capacitance is driven at the sum of the signal and idling frequencies, energy is added to the oscillating system by separating the capacitor plates when the charge is a maximum and is extracted when the charge is zero or a minimum, so that more energy is added than is extracted per cycle of the signal frequency. Hence, gain results. The role of the idling circuit, thus, is to "distort" the charge on the variable capacitor so that energy can be added by the pump. This energy is added in part to the signal and in part to the idling energy so that an enhanced signal at either the signal frequency or the idling frequency can be obtained.

In practice it is impossible to vary a capacitance at the pump frequency required for microwave amplification by mechanical means. Hence pumping is accomplished by using a capacitor whose capacitance varies with voltage, i.e., a nonlinear capacitance, and driving this capacitor with a local oscillator. The nonlinear capacitor is usually used as a coupling element between three resonant circuits, one tuned to the signal frequency, another tuned to the idling frequency, and the third tuned to the pump frequency. The third circuit is driven by the local oscillator, the so-called pump and, because it is resonant at the pump frequency, serves to keep the pump from loading the signal and idling circuits.

DIODE CAPACITORS

It was noted earlier that Torrey and Whitmer studied the use of a crystal diode as a nonlinear capacitance to obtain conversion gain in a mixer. The crystal diode still remains the most attractive means of obtaining a nonlinear capacitance. It exhibits a marked variation of capacitance with applied voltage, it is compact, it can be engineered to fit microwave circuitry and it is cheap, as active microwave devices go.

As is well known, a crystal diode, whether of the point-contact type or of p-n junction type, is a rectifier. It has a current–voltage characteristic as shown in Figure 3. In the forward direction, it conducts

Fig. 3—The current–voltage characteristic of a semiconductor diode.

freely; in the backward direction, the current saturates and remains constant until the reverse voltage reaches approximately 20 volts when avalanche conduction begins. In the saturation region, the conduction current is independent of voltage, so that the diode exhibits an infinite shunt resistance for low-frequency alternating voltages. As the frequency is increased, it becomes apparent that in the saturation region the diode behaves like a capacitor in series with a resistance, the base resistance of the diode. Furthermore, it becomes apparent that the capacitance is not constant but varies with the applied voltage— inversely as the square-root of the voltage for abrupt-junction diodes, inversely as the cube-root of the voltage for graded-junction diodes. Practical diodes usually show a capacitance variation with voltage lying between the inverse square-root and cube-root laws. Figure 4

Fig. 4—The variation of the junction capacitance of a back-biased diode.

shows the variation of capacitance with voltage of a diode designed for use in variable-reactance amplifiers. The capacitance varies almost as the square-root of the voltage and changes by 10 to 1 for a voltage change of about 10 volts.

For purposes of circuit design and for purposes of evaluating diodes for circuit use, it is convenient to represent a diode by an equivalent circuit. Such an equivalent circuit is shown in Figure 5. It consists of a d-c constant-current generator to represent the saturation current, a variable capacitance, a fixed capacitance equal to the lead capacitance or the minimum capacitance available before avalanche conduction sets in, and a resistance to represent the base resistance. The first three of these elements are in parallel and the parallel combination is in series with the resistance. The constant-current generator plays no role in an r-f circuit except as it affects the d-c bias on the diode, and is usually omitted in drawing the equivalent circuit.

$$f_c = \frac{1}{2\pi R_s C_{MIN.}}$$

Fig. 5—Equivalent circuit of a diode.

The equivalent circuit makes it clear what is required of a micro-wave diode capacitor:

1. It should have as rapid a capacitance variation with voltage as possible.
2. C_{min} should be as small as possible so that a substantial voltage can be built up across the capacitance at microwave frequencies.
3. R_s should be as small as possible to keep the losses at a minimum, i.e., to make the electrical Q as high as possible.

The first requirement is circumscribed by the physics of a diode and it may not be possible to do much about it. C_{min} and R_s tend to go together; a reduction in junction area to reduce C_{min} increases R_s correspondingly, so that the Q tends to remain constant.* However,

---

* A commonly used figure of merit for capacitance diodes is the "cut-off frequency" defined by $f_c = 1/(2\pi R_s C_{min})$. It is the frequency at which the diode Q becomes unity.

there are things that can be done, such as reducing the base thickness. The fact that diodes have improved constantly from the early capacitance diodes of Giacoletto and O'Connell[11] to the diode of Uhlir[12] is reassuring for the future.

SIMPLE AMPLIFIER CIRCUITS

As has been noted earlier, the simplest variable-reactance amplifier consists of three tuned circuits coupled together by a nonlinear capacitor. Such an amplifier is shown in Figure 6. The three circuits are tuned, one to the signal frequency, another to the idling frequency, and the third to the pump frequency. Power is fed to the pump circuit by a local oscillator, shown as inductively coupled. Input and output circuits are coupled to the signal circuit. It should be noted that the input and output circuits are effectively in parallel, a matter of some consequence which will be discussed later.

The circuits may take a variety of forms. They may be lumped-constant circuits at low frequencies, they may consist of coaxial lines at microwave frequencies or they may be three independent modes of a single resonant cavity. All have been used. Figure 7 shows a coaxial-line amplifier used by Chang at a 6,600-megacycle signal frequency. It consists of a four-way junction with the pumping signal introduced at the right through a double-stub tuner, with the input and output signals coupled in and out through the double-stub tuner on the left, the idling circuit tuned by the single stub at the rear, and the crystal joining the other three arms from the front. Heffner and Kotzebue have used three independent modes of a single cavity, with the modes coupled to one another by a capacitance diode to achieve an amplifier operating at 2,300 megacycles.[13] The performance of both of these amplifiers is discussed later.

It was pointed out above that the input and output circuits of the simple parametric amplifiers are in parallel, so that these amplifiers are actually two-terminal devices. This is unfortunate because noise originating in the output circuit (in the following amplifier stage, for example) can reach the input, can be amplified and then can return to the output enhanced by the gain of the amplifier. This difficulty can

---

[11] L. J. Giacoletto and J. O'Connell, "A Variable-Capacitance Germanium Junction Diode for UHF," *RCA Review*, Vol. XVII, p. 68, March, 1956.

[12] A. Uhlir, Jr., "The Potential of Semiconductor Diodes in High-Frequency Communications," *Proc. I.R.E.*, Vol. 46, p. 1099, June, 1958.

[13] H. Heffner and K. Kotzebue, "Experimental Characteristics of a Microwave Parametric Amplifier Using a Semiconductor Diode," *Proc. I.R.E.*, Vol. 46, p. 1301, June, 1958.

Fig. 6—The circuit of a simple parametric amplifier.

be circumvented by using a circulator to isolate the input and output circuits. The arrangement is shown in Figure 8. The input signal enters the circuit at the left and proceeds counter-clockwise through one quadrant to the amplifier. The output signal proceeds through the next quadrant to the following stage, the receiver in the figure. Any noise generated in the input of the receiver proceeds counter-clockwise through the third quadrant to a terminating resistor. Thus, the input and output circuits are isolated and noise cannot be fed back into the amplifier from the following amplifier.

While circulators afford an attractive method of separating the input and output circuits of two-terminal devices, they are not undisguised blessings. They are themselves responsible for noise and may very well contribute more noise to the output than the amplifier itself. Furthermore, they are bulky and in some instances contribute 80 to 90 per cent of the volume and weight of a parametric amplifier. These considerations have led to a search for four-terminal parametric amplifiers, several of which are discussed in the next section.

Fig. 7—Photograph of a 6,600-megacycle parametric amplifier.

MORE-SOPHISTICATED AMPLIFIERS

The noise figure of a parametric amplifier depends on the ratio of the signal frequency to the idling frequency, the smaller the ratio the lower the noise figure. This consideration has led to an investigation of up-converters—amplifiers in which the pump frequency is made high compared to the signal frequency and input–output isolation is achieved by taking the output at the idling frequency.[14,15] Such amplifiers achieve a very low noise figure.

Up-converters are very attractive for lower frequencies. However, at X-band or higher frequencies they pose problems in finding suitable local oscillators and circuits to handle the output signal. The problem

Fig. 8—Parametric amplifier with circulator to separate input and output.

of finding pumping sources suitable for future millimeter-wave amplifiers prompted Chang and Bloom to investigate amplifiers with the pump at a frequency lower than the signal frequency.[16,17] In this mode of operation, they in effect use the nonlinearity of the capacitor diode to generate a second or third harmonic of the pump frequency which interacts with the signal to achieve gain. Amplifiers operating at 200, 380, and 6,600 megacycles using lower-frequency pumping have been demonstrated.

To eliminate the need for a circulator in an amplifier with the input and output frequencies equal, Chang has devised a four-terminal amplifier using three stages.[18] The arrangement is shown in Figure 9. It

[14] G. F. Hermann, M. Uenohara and A. Uhlir, Jr., "Noise Figure Measurements on Two Types of Variable Reactance Amplifiers Using Semiconductor Diodes," *Proc. I.R.E.*, Vol. 46, p. 1301, June, 1958.

[15] B. Salzberg and E. W. Sard, "A Low Noise, Wide-Band Reactance Amplifier," *Proc. I.R.E.*, Vol. 46, p. 1303, June, 1958.

[16] S. Bloom and K. K. N. Chang, "Parametric Amplifiers Using Lower-Frequency Pumping," *Jour. Appl. Phys.*, Vol. 29, p. 594, March, 1958.

[17] K. K. N. Chang and S. Bloom, "A Padametric Amplifier Using Lower-Frequency Pumping," *Proc. I.R.E.*, Vol. 46, p. 1383, July, 1958.

[18] K. K. N. Chang, "A Four-Terminal Parametric Amplifier," *Proc. I.R.E.*, Vol. 47, p. 81, January, 1959.

Fig. 9—Chang's four-terminal parametric amplifier.

consists of a down-converter followed by an amplifier at the interme-
diate frequency, followed in turn by an up-converter which restores
the original frequency. All of the three stages are pumped from a
common source which, in Chang's reduction to practice, is at a fre-
quency lower than the signal frequency. If the input and output were
pumped equally, the network would be reciprocal. However, Chang
proportions the pumping powers to the three stages to achieve a
50-decibel input–output isolation. By using down-conversion in the
first stage, he makes the ratio of signal frequency to idling frequency
small in the second stage, a situation favorable for a low-noise figure.

Thus far, nothing has been said about the bandwidth of variable-
reactance amplifiers. Because they are essentially regenerative, these
amplifiers are subject to the same bandwidth limitations as other
regenerative amplifiers. The bandwidth situation may be described
as follows: Given a passive network with a prescribed bandwidth,
the gain obtainable in the circuit by regeneration is proportional to
the reduction in bandwidth tolerable for the application. Hence, if the
bandwidth of a variable-reactance amplifier with the pump turned off
is just equal to the bandwidth required for a given application, the
bandwidth will be too narrow when the pump is turned on to achieve
gain.

$$f_p = f_s + f_i$$

$$f_p < f_c < f_p + f_s$$

$$f_c = \frac{1}{2\pi\sqrt{L\bar{C}}}$$

Fig. 10—Schematic diagram of a traveling-wave parametric amplifier.

This consideration has led to an investigation of circuits with very wide passive bandwidths. Such a circuit is the low-pass filter. This may be made into a traveling-wave variable-reactance amplifier by replacing the usual shunt capacitors by capacitance diodes as shown in Figure 10. The signal and the pumping power are introduced at the left-hand end of the network. The right-hand end is terminated so that no reflection occurs. There is then no feed-back, and the amplifier becomes a four-terminal device with input and output separated. In order to achieve amplification, the signal frequency, the pump frequency, and the idling frequency must be in the pass-band of the network, but the upper side band—the signal having a frequency equal to the sum of the signal frequency and the pump frequency—must be in the stop band.

The general theory of such traveling-wave amplifiers has been developed by Tien[19] and an amplifier operating on these principles has been described by Engelbrecht.[20]

PERFORMANCE DATA

Now that the various variable-reactance amplifiers using nonlinear capacitances have been described, it is pertinent to examine the performance they yield. The experimental performance data are summarized in Figure 11. The data are self-explanatory and only a few comments are needed:

Cavity amplifier (Row 2): The two cases shown differ only in the relation of the signal frequency to the pump frequency. The second case has the signal at just one-half of the pump frequency so that the signal and idling frequencies are equal. This situation leads to a lower noise figure than does the case where the signal and idling frequencies differ, because the noise bandwidth coincides with the signal bandwidth and there is no idling circuit noise.

Coaxial amplifier (Row 3): Note the reduction in bandwidth as the pumping power is increased to increase the gain. The product of voltage gain and bandwidth is a constant.[17] Also note the pumping power with the low-frequency pump in mind.

Up-converters (Rows 5 and 6): The up-converters show the best noise figures. These results accord with the theoretical predictions that a small ratio of signal frequency to idling frequency is favorable for a low noise figure. This favorable situation also obtained in the second

[19] P. K. Tien, "Parametric Amplification and Frequency Mixing in Propagating Circuits," *Jour. Appl. Phys.*, Vol. 29, p. 1347, September, 1958.

[20] R. S. Engelbrecht, "A Low-Noise Non-Linear Reactance Traveling-Wave Amplifier," *Proc. I.R.E.*, Vol. 46, p. 1655, September, 1958.

TYPE	f_{IN} MC	f_p MC	f_{OUT} MC	P_p mw	G db	Δf MC	NF db	REF.
CAVITY	2300	3500	2300	$<10^3$	19	1	< 4.8	13
CAVITY	5840	11,700	5840	50-500	18	8	< 6	14
	5850	11,700	5850	50-500	~18	~8	3	14
COAXIAL	380	300	380	< 30	10	3.4	7	17
	380	300	380	30	30	0.38	8	17
4-TERMINAL	214	150	214	100	8	0.25	2.5	18
UP-CONVERTER	460	8915	9375	200	9 ✱		2±0.5 ✱	14
	460	8915	9375	200	21 ✱		1.1±0.5 ✱	14
UP-CONVERTER	1	20	21	—	10	0.1	0.4	15
T W AMP.	380	630	380	2	10-12	10-20	3.5	20

✱ A PARTICULAR ADJUSTMENT, NOT REPRODUCIBLE AT WILL.

Fig. 11—Reported performance data of some parametric amplifiers. f_{in} = input frequency, f_p = pump frequency, f_{out} = output frequency, P_p = pumping power, G = power gain, f = bandwidth, NF = noise factor. f_{in} and f_{out} are the same except in the case of the up-converters.

Fig. 12—The status of parametric amplifiers relative to other microwave amplifiers.

stage of the four-terminal amplifier (Row 4) and accounts for its noise figure.

Traveling-wave amplifier (Row 7): The bandwidth is noteworthy. The circuits with wide passive bandwidths are showing their merit.

The salient features of these experimental results seem to the writer to be:

1. Parametric amplifiers are beginning to show the low noise figures predicted for them.

2. It is possible to achieve good performance with the pump at a frequency lower than the signal frequency.

3. Wide-band parametric amplifiers are achieveable.

CONCLUSION

In concluding this discussion of nonlinear-capacitance amplifiers, it is of interest to examine the status of present amplifiers of this kind with respect to other amplifiers in noise figure and their status with respect to possible applications. The situation is summarized in Figure 12. The figure shows the equivalent noise temperatures of terrestrial and extraterrestrial noise sources and the noise temperatures of various microwave amplifiers, all versus frequency. The noise temperatures of galactic noise sources appear at the left, and decrease rapidly with frequency. At about 500 megacycles, galactic sources have about the same temperature as the noise temperature of the earth (about 300°K). Above 1,000 megacycles, noise from galactic sources (aside from radio stars) becomes negligible. Above 10,000 megacycles atmospheric noise becomes of consequence and rises to about the temperature of the earth at 100,000 megacycles. The dips are "windows," the first one the familiar window at 8.6 millimeters wavelength.

The three noise sources, galactic sources, the atmosphere and the earth, determine what is required of an amplifier. If the antenna looks upward and is highly directional so that it does not see the earth through spill-over or side lobes, the amplifier should have a noise temperature as low as the galactic noise temperature if the ultimate in performance is to be obtained. For antennas which see the earth, amplifiers with equivalent noise temperatures below that of the earth are adequate.

The equivalent noise temperatures of some familiar triodes are shown at the left in Figure 12. The 416B appears quite adequate for frequencies up to about 500 megacycles. Crystal mixers are moderately good up to about 10,000 megacycles in applications where the antenna sees the earth. Commercial low-noise traveling-wave tubes match the

performance of crystal mixers at 3,000 megacycles and laboratory tubes exceed crystal performance. For upward-looking antennas, no amplifier matches the maser in performance.

The crosses in Figure 12 show the noise temperatures of the variable-reactance amplifiers discussed above. The amplifiers operating at frequencies below 1,000 megacycles show noise temperatures adequate for any application. The one amplifier operating at a frequency above 1,000 megacycles has a noise figure better than the best traveling-wave tube. However, it has a long way to go before it competes with masers in upward-looking applications. Lest present performance sound discouraging, it should be pointed out that all of the variable-reactance amplifiers discussed have operated at room temperature and that the noise figure of a variable-reactance amplifier can be reduced by cooling in the same manner that it is reduced in masers. With this in mind, it seems to the writer that the nonlinear-capacitance amplifier has come a long way in a brief time, and that before long, this amplifier with its simplicity, compactness, and low-noise performance, will find many applications, both military and civilian.

THE EXPONENTIAL GUN—A LOW-NOISE GUN FOR TRAVELING-WAVE-TUBE AMPLIFIERS*

By

A. L. Eichenbaum and R. W. Peter

RCA Laboratories,
Princeton, N. J.

Summary—In every low-noise beam-type amplifier, three major regions can be distinguished: the cathode region in front of the emitter where fluctuations of the electron velocity are comparable to their average velocity, the transformer or "gun" region between the cathode region and the circuit input, and the interaction region between the circuit input and output.

This paper deals with the "gun" region only. This region is treated as a transmission-line matching transformer. It is assumed that the conditions at both ends of the transformer are given. Consideration of the gun requirements for a low-noise traveling-wave tube leads to the exponential gun as the best presently known solution.

The exponential gun can satisfy the low-noise boundary conditions over a broad frequency band. It is free of excess noise created by beam-voltage discontinuities, since in the exponential gun the voltage along the beam increases approximately linearly. In addition, it is demonstrated that with an exponential gun one can minimize the noise not only for the fundamental space-charge wave mode, but also for several higher-order modes.

INTRODUCTION

THE problem of low-noise amplification in a traveling-wave tube is very complex.[1-5] In simplified terms, it may be represented as follows:

At the cathode, electrons are emitted with random velocity and in random numbers. They cause a whole spectrum of space-charge

---

* Manuscript received November 13, 1958.

[1] F. N. H. Robinson, "Microwave Shot Noise in Electron Beams and the Minimum Noise Factor of Traveling Wave Tubes and Klystrons," *Journal British I.R.E.*, Vol. 14, p. 79, February, 1954.

[2] S. Bloom and R. W. Peter, "A Minimum Noise Figure for the Traveling Wave Tube," *RCA Review*, Vol. XV, p. 252, June, 1954.

[3] J. R. Pierce and W. E. Danielson, "Minimum Noise Figure of Traveling Wave Tubes with Uniform Helices," *Jour. Appl. Phys.*, Vol. 25, p. 1163, September, 1954.

[4] H. A. Haus, "Noise in One-Dimensional Electron Beams," *Jour. Appl. Phys.*, Vol. 26, p. 560, May, 1955.

[5] S. Bloom, "Effect of Initial Noise Current and Velocity Correlation on the Noise Figure of Traveling Wave Tubes," *RCA Review*, Vol. XVI, p. 179, June, 1955.

waves to be excited in the beam. These noise waves propagate along the beam through the gun and are transformed by it before the beam enters the interaction region at the circuit input. At this point, the signal is introduced and amplified together with the noise in the beam. One can thus discern three main regions:

(1) The cathode region, the short region ahead of the emitting surface, where the fluctuations of the electron velocity are comparable in order of magnitude to the average velocity.

(2) The transformer or gun region extending from the end of the cathode region to the circuit input plane. In this region the velocity fluctuations of the electrons are small compared to the average velocity.

(3) The interaction region which extends between the circuit input and output planes.

Regions 1 and 2 together constitute the gun. This paper deals with the second region of the gun, the transformer region, only.*

According to previous work, the propagation of waves along an electron beam can be analyzed analogously to the propagation of electromagnetic waves along a transmission line.[6] Assuming that the boundary conditions at the circuit input plane and at the end of the cathode region are given, the design of the transforming region is reduced to a problem of optimizing a transmission-line transformer which satisfies the given boundary conditions. This study, therefore, has been isolated from noise considerations as such; noise considerations are implicitly contained in the boundary conditions assumed.

The required boundary conditions at the helix input plane for minimum noise factor were described by Bloom and Peter[2] for the case of uncorrelated noise excitation. Later Haus[4] and Bloom[5] described the conditions for the more general case of partial correlation between current and velocity excitation. These conditions were derived for the fundamental space-charge wave and were found to be unique and independent of the degree of correlation. The more complex problem of including the higher harmonic modes has also been investigated, but a complete solution is not yet available. The situation in the

---

* In the following, the term "low-noise gun" refers to this region.

[6] S. Bloom and R. W. Peter, "Transmission-Line Analog of a Modulated Electron Beam," *RCA Review*, Vol. XV, p. 95, March, 1954.

cathode region is presently an object of investigation[7-10] and much improvement in this area may be expected.

Regardless of the detailed results which will emerge from further studies of regions 1 and 3, the problem of optimizing the transformer region 2 can be isolated and considered separately, once its boundary conditions are given.

The two basic assumptions are used throughout this paper, in discussing the transformer region;

(1) The transmission-line analog of space-charge-wave propagation along an electron beam is valid. This requires that the a-c velocity fluctuations in the beam be considerably smaller than the d-c beam velocity (this condition holds by definition for region 2).

(2) The beam behaves as a linear wave-transmission medium. This means that the different modes of propagation are assumed to propagate independently without mutual coupling. This assumption has been shown to hold exactly for a drifting beam surrounded by a shield, both of which have a constant cross-section,[11] and approximately for a beam of arbitrary d-c velocity if surrounded tightly by a shield.[12]

The present paper first presents previous work on space-charge-wave transformers in a unified transmission-line formalism. Then the exponential tranformer is singled out as the preferred solution for a traveling-wave tube low-noise gun, and the three-region gun is discussed as a practical embodiment of the exponential gun principle. Finally a solution is presented for multimode matching by use of an exponential gun.

The work described in this paper was done during the years 1954 and 1955. A part of it was included in a paper given at the Electron

[7] A. E. Siegman and S. Bloom, "An Equivalent Circuit for Microwave Noise at the Potential Minimum," *Trans. I.R.E. PGED*, October, 1957.

[8] J. Berghammer, "Space Charge Waves in Hydrodynamic Model of a Multivelocity Electron Beam," The 16th Annual Conference on Electron Tube Research, University Laval, Quebec, Canada, 1958.

[9] W. E. Vivian, "Transport of Cathode Shot Noise Through a Space Charge Limited Diode at Microwave Frequencies," The 16th Annual Conference on Electron Tube Research, Quebec, Canada, 1958.

[10] J. Feinstein, "The Noise Parameters of a Multivelocity Electron Stream," The 16th Annual Conference of Electron Tube Research, Quebec, Canada, 1958.

[11] S. Ramo, "The Electron-Wave Theory of Velocity-Modulated Tubes," *Proc. I.R.E.*, Vol. 27, p. 757, December, 1939.

[12] Unpublished work of S. Bloom.

Tube Research Conference in June, 1955, and also in a summer course on noise at Massachusetts Institute of Technology in August, 1955.[13]

FUNDAMENTAL SPACE-CHARGE-WAVE TRANSFORMERS

The analogy between the differential equations for propagation of space-charge-waves along an electron beam (for which our two basic assumptions hold) and for propagation of electromagnetic waves along a lossless transmission line[6] permits one to use the formalism of transmission-line theory in the description of a low-noise gun. The wave-transducing qualities of any section of an electron beam are, therefore,

$$V_{i+1} = A_i V_i + B_i I_i,$$

$$I_{i+1} = C_i V_i + D_i I_i.$$

In the form of a chain matrix of a purely reactive, passive four-terminal network, this is written

$$\left\| \begin{array}{c} V_{i+1} \\ I_{i+1} \end{array} \right\| = \left\| \begin{array}{cc} A_i & B_i \\ C_i & D_i \end{array} \right\| \times \left\| \begin{array}{c} V_i \\ I_i \end{array} \right\| \tag{1}$$

Using the same terminology and notation as in Reference (6), I and V represent the r-m-s values of the beam current and of the kinetic potential fluctuations, respectively. Since the transducer is passive, its matrix determinant vanishes, i.e., $A_i D_i - B_i C_i = 0$. As it is also loss-free, A_i and D_i are real, while B_i and C_i are imaginary. For every space-charge wave mode, this matrix is different.

Any low-noise gun may be subdivided into a chain of n elementary transformer sections, each with its transformer matrix (M_i). If the matrix of each section is known, the matrix of the complete gun can be represented by

$$\left\| \begin{array}{c} V_n \\ I_n \end{array} \right\| = \left\| M_{n-1} \times M_{n-2} \times \cdots M_i \times \cdots \cdots M_1 \right\| \times \left\| \begin{array}{c} V_1 \\ I_1 \end{array} \right\| \tag{2}$$

Every transformer (M_i) transforms the impedance $Z_i = V_i/I_i$ at plane i into Z_{i+1} at plane $(i + 1)$ according to Equation (1).

$$Z_{i+1} = \frac{V_{i+1}}{I_{i+1}} = \frac{A_i Z_i + B_i}{C_i Z_i + D_i}. \tag{3}$$

---

[13] R. W. Peter, "Noise in Electron Devices," Lectures delivered at the Massachusetts Institute of Technology, August, 1955.

Conveniently, the beam impedance is normalized with respect to the characteristic beam impedance W, i.e., $z = Z/W$.

For computing and designing of a low-noise gun, it is advantageous to use the most simple fundamental space-charge-wave transformers (M_i) as building elements. These will be discussed in the following section.

General Transmission-Line Equations

The space-charge-wave transforming properties of an electron beam can be described by the following transmission-line equations.[6]

$$\frac{dV}{d\phi} = jWI,$$

$$\frac{dI}{d\phi} = j\frac{V}{W}, \tag{4}$$

where W is the characteristic beam impedance. By a differentiation of Equation (4) with respect to ϕ, one obtains

$$\frac{d^2V}{d\phi^2} - \left(\frac{1}{W}\frac{dW}{d\phi}\right)\frac{dV}{d\phi} + V = 0,$$

$$\frac{d^2I}{d\phi^2} + \left(\frac{1}{W}\frac{dW}{d\phi}\right)\frac{dI}{d\phi} + I = 0. \tag{5}$$

This is a pair of general linear differential equations if $(dW/d\phi)/W$ is a function of ϕ. The general Equations (5) can be solved by the method of G. Frobenius.[14]

Four Special Solutions:

The general space-charge-wave Equations (5) can be solved explicitly for a few simple functions $(dW/d\phi)/W$. Solutions for four special cases will be discussed. They represent the most basic space-charge-wave transformers:

(a)   The drifting beam,[11,15]

(b)   The "Bessel" transformer,[16-18]

[14] E. T. Whittaker and G. N. Watson, *Modern Analysis*, Cambridge University Press, Cambridge, England, 1927.
[15] W. C. Hahn, "Small Signal Theory of Velocity-Modulated Electron Beams," *General Electric Review*, Vol. 42, p. 258, June, 1939.
[16] P. K. Tien and L. M. Field, "Space-Charge Waves in an Accelerated Electron Stream for Amplification of Microwave Signals," *Proc. I.R.E.*, Vol. 40, p. 688, June, 1952.

(c)   The parallel-plane diode,[19,20]

(d)   The exponential transformer.

The first three of these transformers have been discussed in the literature, as indicated by references given.

(a)   The Drifting Beam

In this case, the characteristic beam impedance W is constant; therefore,

$$\frac{1}{W}\frac{dW}{d\phi} = 0. \tag{6}$$

Hence, Equations (5) become

$$\frac{d^2V}{d\phi^2} + V = 0,$$

$$\frac{d^2I}{d\phi^2} + I = 0. \tag{7}$$

With the following initial conditions at $\phi = \phi_0$:

$$V = V_1, \qquad \frac{dV}{d\phi} = jW_1I_1,$$

$$I = I_1, \qquad \frac{dI}{d\phi} = j\frac{V_1}{W_1}, \tag{8}$$

we find

$$\left\| \begin{array}{c} V \\ I \end{array} \right\| = \left\| \begin{array}{cc} \cos\phi & jW_1\sin\phi \\ \dfrac{j}{W_1}\sin\phi & \cos\phi \end{array} \right\| \times \left\| \begin{array}{c} V_1 \\ I_1 \end{array} \right\|, \tag{9}$$

which is the well-known matrix description for the uniform transmission line.

[17] R. Mueller, "Raumladungswellen in Elektronenstroemungen," *Archiv. Elektrischen Ubertragung*, Vol. 9, p. 505, November, 1955.

[18] Nasaaki Higuchi, "Propagation of Space-Charge Waves in an Accelerated Electron Beam," Report of the Microwave Communication Res. Committee in Japan, p. 87, December, 1955.

[19] F. B. Llewellyn and L. C. Peterson, "Vacuum Tube Networks," *Proc. I.R.E.*, Vol. 32. p. 144, March, 1944.

[20] L. D. Smullin, "Propagation of Disturbances in One-Dimensional Accelerated Electron Streams," *Jour. Appl. Phys.*, Vol. 22, p. 1495, December, 1951.

(b)   The Bessel-Type Transformer

Tien et al.,[16] Mueller,[17] and Higuchi[18] have treated the case of Bessel-type transformers. In transmission line analogy, a Bessel-type transformation is obtained by stipulating a characteristic beam impedance variation,

$$W = W_1 \phi^a, \tag{10}$$

that is

$$\frac{1}{W} \frac{dW}{d\phi} = \frac{a}{\phi}. \tag{11}$$

The constant a determines the steepness of the impedance rise along the beam. Substituting Equation (11) into Equations (5) leads to a pair of Bessel-type differential equations;

$$\frac{d^2V}{d\phi^2} - \frac{a}{\phi} \frac{dV}{d\phi} + V = 0,$$

$$\frac{d^2I}{d\phi^2} + \frac{a}{\phi} \frac{dI}{d\phi} + I = 0. \tag{12}$$

The solution to Equations (12) can be found in Jahnke and Emde.[21] Letting

$$\frac{1+a}{2} = u + 1 \tag{13}$$

and using the four initial conditions given by Equations (8), the solution to Equations (12) can be written as

$$\left\| \begin{array}{c} V \\ I \end{array} \right\| = \left\| \begin{array}{cc} A & B \\ C & D \end{array} \right\| \times \left\| \begin{array}{c} V_1 \\ I_1 \end{array} \right\|, \tag{14}$$

where

$$A = \left(\frac{\phi}{\phi_0} \right)^{u+1} \frac{\pi\phi_0}{2} [Y_u(\phi_0) J_{u+1}(\phi) - J_u(\phi_0) Y_{u+1}(\phi)],$$

$$B = jW_1(\phi\phi_0)^{u+1} \frac{\pi}{2} [J_{u+1}(\phi_0) Y_{u+1}(\phi) - Y_{u+1}(\phi_0) J_{u+1}(\phi)], \tag{15}$$

---

[21] E. Jahnke and F. Emde, *Tables of Functions*, Dover Publications, Inc., New York, 1945, p. 146.

$$C = j \frac{1}{W_1(\phi\phi_0)^u} \frac{\pi}{2} [J_u(\phi_0)Y_u(\phi) - Y_u(\phi_0)J_u(\phi)],$$

$$D = \left(\frac{\phi_0}{\phi}\right)^u \frac{\pi\phi_0}{2} [J_{u+1}(\phi_0)Y_u(\phi) - Y_{u+1}(\phi_0)J_u(\phi)].$$

(c)   The Parallel Plane Diode

Propagation along a parallel electron flow of infinite transverse extent has been studied by Llewelyn,[19] Smullin,[20] and others. Using the transmission-line formalism, the characteristic beam impedance of a space-charge-limited diode transformer has been given by Bloom and Peter.[6]

Expressed as a function of ϕ, it becomes

$$W = W_1 \cosh^3 \frac{\phi}{\sqrt{2}}, \tag{16}$$

and consequently

$$\frac{1}{W} \frac{dW}{d\phi} = \frac{3}{\sqrt{2}} \tanh \frac{\phi}{\sqrt{2}}. \tag{17}$$

Substituting $(dW/d\phi)/W$ into Equation (5) leads to a pair of general linear differential equations. Using initial condition of Equations (8), the solution for the planar space-charge-limited diode becomes*

$$\begin{Vmatrix} V \\ \\ I \end{Vmatrix} = \begin{Vmatrix} \left(1 - \sinh^2 \frac{\phi}{\sqrt{2}}\right) & jW_1\sqrt{2}\sinh\frac{\phi}{\sqrt{2}} \\ \\ \dfrac{j\sqrt{2}\sinh\frac{\phi}{\sqrt{2}}}{W_1\cosh^2\frac{\phi}{\sqrt{2}}} & \dfrac{1}{\cosh^2\frac{\phi}{\sqrt{2}}} \end{Vmatrix} \times \begin{Vmatrix} V_1 \\ \\ I_1 \end{Vmatrix} \tag{18}$$

The corresponding normalized impedance, obtained from Equations (3) and (18) is

---

* This solution is identical with the one obtained by Bloom and Peter[6] if their variable τ is expressed in terms of ϕ. The variation of ϕ with τ or with V_0/V_{0_1} may be obtained from Figure 1.

Fig. 1—Phase angle, ϕ, in plasma-frequency radians in a space-charge-limited planar diode.

Fig. 2—Impedance transformation of space-charge-limited planar diode for initial normalized impedance $r = 1$, 1.4, 2, 3.

$$z = \frac{\left[1 - \sinh^2 \dfrac{\phi}{\sqrt{2}} \right] z_1 + j\sqrt{2} \sinh \dfrac{\phi}{\sqrt{2}}}{\cosh \dfrac{\phi}{\sqrt{2}} \left[j\sqrt{2} \sinh \dfrac{\phi}{\sqrt{2}} z_1 + 1 \right]}, \tag{19}$$

with the initial beam impedance $Z_1 = V_1/I_1$ and $z_1 = Z_1/W_1$ in normalized form. The normalized impedance, Equation (19), has been

Fig. 3—Impedance transformation of space-charge-limited planar diode for initial normalized impedance $r = 1$, 0.5, 0.3, 0.2.

plotted on a Smith chart in Figure 2 for values of normalized initial impedance $z_1 = r_1 = 1, 1.4, 2, 3$, and in Figure 3 for values $z_1 = r_1 = .2$, .3, .5 and 1. Lines of equal phase angle are shown in both figures. It may be noted that for small values of ϕ, a diode behaves like a constant-impedance transmission line which simply shifts the phase angle but does not change the standing wave ratio.

(d)   The Exponential Transformer

This transformer is another one for which the transmission-line Equations (5) can be solved explicitly. This transformer represents the beam equivalent of the exponential acoustic horn or exponential transmission line often used for impedance matching, for which the characteristic impedance varies as

$$W = W_1 e^{k\phi}, \tag{20}$$

where the rate of impedance change, k, is a constant,

$$\frac{1}{W}\frac{dW}{d\phi} = k. \tag{21}$$

W_1 is the initial characteristic impedance. Substituting Equation (21) into the general transmission line Equations (5),

$$\frac{d^2V}{d\phi^2} - k\frac{dV}{d\phi} + V = 0, \tag{22}$$

$$\frac{d^2I}{d\phi^2} + k\frac{dI}{d\phi} + I = 0. \tag{23}$$

Letting

$$m = \left| \sqrt{\left(\frac{k}{2}\right)^2 - 1} \right| \tag{24}$$

and using initial conditions (Equation (8)), solutions to the linear differential Equations (22) and (23) are as follows:

For the case $|k| > 2$, the solution to Equations (22) and (23) is

$$\left\| \begin{array}{c} V \\ I \end{array} \right\| = \left\| \begin{array}{cc} e^{k\phi/2}\left(\cosh m\phi - \dfrac{k}{2m}\sinh m\phi\right) & je^{k\phi/2}\dfrac{W_1}{m}\sinh m\phi \\ \dfrac{je^{-k\phi/2}}{W_1 m}\sinh m\phi & e^{-k\phi/2}\left(\cosh m\phi + \dfrac{k}{2m}\sinh m\phi\right) \end{array} \right\| \times \left\| \begin{array}{c} V_1 \\ I_1 \end{array} \right\| \tag{25}$$

For the case $k < 2$, the solution to Equations (22) and (23) are periodic; $\cosh m\phi$ and $\sinh m\phi$ in Equation (25) are replaced by $\cos m\phi$ and $\sin m\phi$ respectively. In the special case of $k = 2$, Equation (25) reduces to

$$\left\| \begin{array}{c} V \\ I \end{array} \right\| = \left\| \begin{array}{cc} e^{k\phi/2}\,(1 - k\phi/2) & je^{k\phi/2}\,W_1\phi \\ je^{-k\phi/2}\,\dfrac{\phi}{W_1} & e^{-k\phi/2}\,(1 + k\phi/2) \end{array} \right\| \times \left\| \begin{array}{c} V_1 \\ I_1 \end{array} \right\| \tag{26}$$

The normalized impedance of the exponential transformer for $k > 2$ as obtained with Equations (3) and (25) is

$$z(m\phi) = \frac{-2z_1 m^2 + j\left\{ (z_1{}^2 - 1)\, m \sinh 2m\phi \; - (1 + z_1{}^2)\,\dfrac{k}{2} \cosh 2m\phi + (1 + z_1{}^2)\,\dfrac{k}{2} \right\}}{1 + z_1{}^2 + \left(1 - \dfrac{k^2}{2} - z_1{}^2 \right) \cosh 2m\phi - km \sinh 2m\phi}. \tag{27}$$

Figures 4, 5, and 6 are plots of $z(m\phi)$ on a Smith chart for $k = 2.12$, 3, and 4. The exponential transformer with $k > 2$ behaves for all frequencies like a transmission line below cutoff. If $m\phi$ goes to infinity, the input impedance becomes purely imaginary. Comparison of Figures 2 and 4 shows that an exponential transformer with $k = 2.12$ closely approximates the plane parallel diode at large angles ϕ. The solid equi-phase-angle curves in Figures 4, 5, and 6 indicate angles $m\phi$ for $z(m\phi)$ curves starting at the right half of the real axis, i.e., for impedance z_1 real and larger than unity. The dotted equi-phase-angle lines indicate angles $m\phi$ for $z(m\phi)$ curves starting at the left half of the real axis, i.e., for terminating impedance z_1 real and smaller than unity.

For the case of $k < 2$, with $\cos m\phi$ and $\sin m\phi$ in Equation (25),

$$z(m\phi) = \frac{2z_1 m^2 + j\left\{ (1 - z_1{}^2)\, m \sin 2m\phi \; - (1 + z_1{}^2)\,\dfrac{k}{2} \cos 2m\phi + (1 + z_1{}^2)\,\dfrac{k}{2} \right\}}{1 + z_1{}^2 + \left(1 - \dfrac{k^2}{2} - z_1{}^2 \right) \cos 2m\phi + km \sin 2m\phi}; \tag{28}$$

$z(m\phi)$ is periodic in $m\phi$. Figures 7 and 8 are families of impedance transformation circles for values of $k = 0.4$ and $k = 1$, with z_1 as the parameter. The center of the family of circles depends on the value of k. The input-impedance circle for an exponential line with imperfect termination and $k < 2$ surrounds the circle for perfect match. The diameter of the $z(m\phi)$ circle depends on the degree of mismatch and on the steepness of impedance rise, k. For large mismatch values and

a fast impedance rise the diameter of the $z(m\phi)$ circle is large. For the case of perfect match, $z_1 = 1$, the maximum reflection coefficient is equal to the diameter of the circle, which is $k/2$.

The mismatches $z_1 = r_1$ and $z_1 = 1/r_1$ yield identical z circles due to the symmetry and periodicity of the solution.

The normalized impedance for the case of perfect termination, $z_1 = 1$, is obtained from Equations (27) and (28). For $k > 2$,

Fig. 4—Normalized impedance $z = Z/W$ of an exponential transformer with characteristic impedance $W = W_1 e^{k\phi}$ for $k = 2.12$. Curves are given for initial normalized impedances $z_1 = r_1 = 4$, 3, 2, 1, ½, ⅓, and ¼. Solid $m\phi$ lines are for terminating impedances $z_1 = r_1 = 4$, 3, 2, and 1. Dotted $m\phi$ lines are for terminating impedances $z_1 = r_1 = 1$, ½, ⅓, and ¼.

$$z = \frac{-m^2 - jk \sinh^2 m\phi}{1 - \dfrac{k}{2} \cosh (2m\phi + \alpha)}, \tag{29}$$

where

$$\alpha = \tanh^{-1} \frac{2m}{k}, \tag{30}$$

and for $k < 2$,

$$z_1 = \frac{m^2 + jk \sin^2 m\phi}{\dfrac{k}{1 + \dfrac{k}{2} \cos (2m\phi + \alpha')}}, \tag{31}$$

where

$$\alpha' = \tan^{-1} \frac{2m}{k}. \tag{32}$$
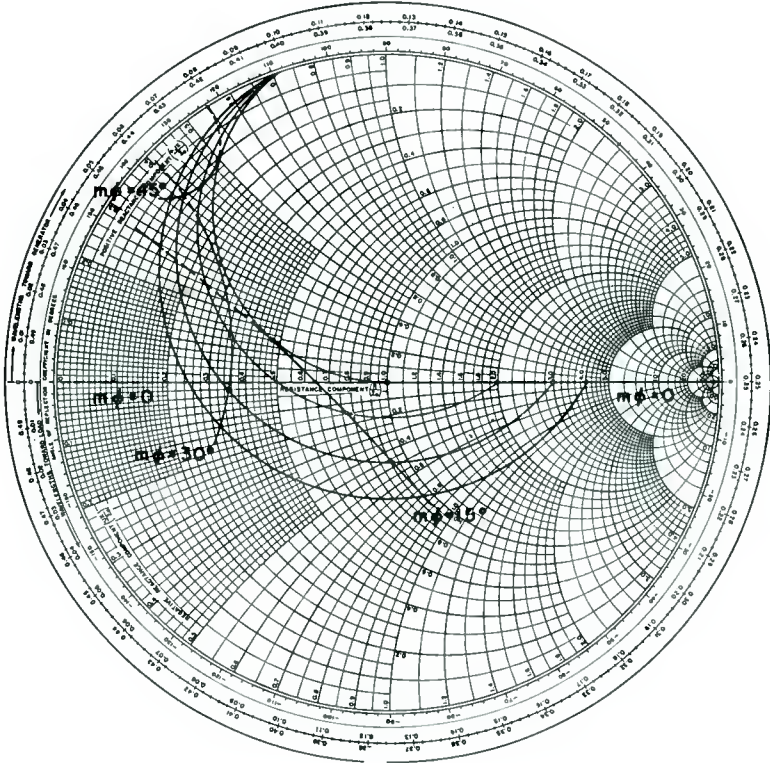

Fig. 5—Normalized impedance $z = Z/W$ of an exponential transformer with characteristic impedance $W = W_1 e^{k\phi}$ for $k = 3$. Curves are given for initial normalized impedances $z_1 = r_1 = 4$, 3, 2, 1, ½, ⅓, and ¼. Solid $m\phi$ lines are for terminating impedances $z_1 = r_1 = 4$, 3, 2, and 1. Dotted $m\phi$ lines are for terminating impedances $z_1 = r_1 = 1$, ½, ⅓, and ¼.

A plot of $z(m\phi)$ for values of $k < 2$, $k = 2$ and $k > 2$ is shown in Figure 9 for the case of perfect termination $(z_1 = r_1 = 1)$. Values of $k > 2$ do not give periodic solutions for $z(m\phi)$. As before, equiphase curves $(m\phi = \text{constant})$ are shown with light lines. For $k \to 0$, the drifting beam solution, Equation (9), is approached.

Potential Distribution in Space-Charge Wave Transformers

With a characteristic beam impedance, W, specified as a function of phase angle ϕ, the corresponding potential distribution along the
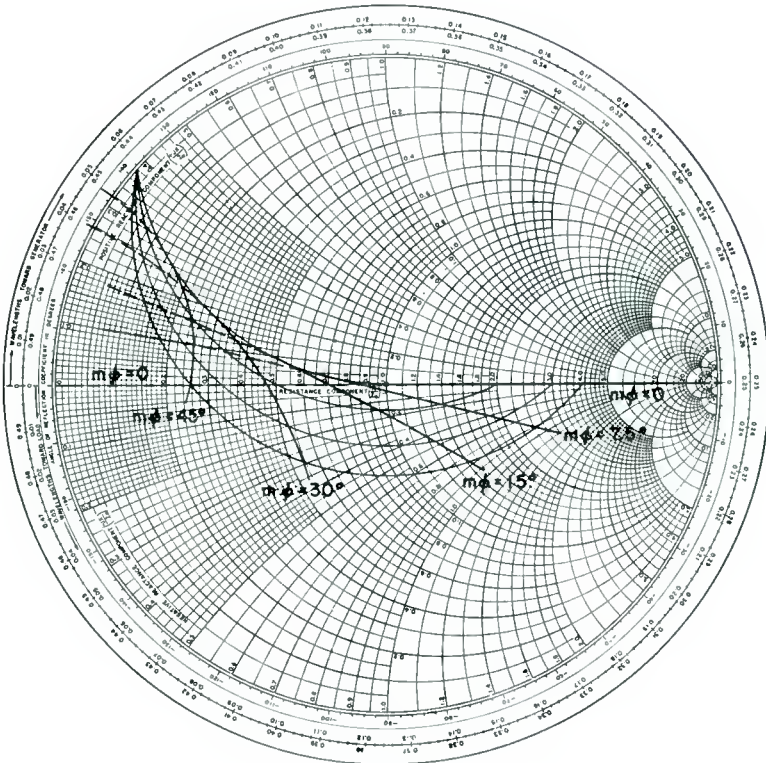

Fig. 6—Normalized impedance $z = Z/W$ of an exponential transformer with characteristic impedance $W = W_1 e^{k\phi}$ for $k = 4$. Curves are given for initial normalized impedances $z_1 = r_1 = 4$, 3, 2, 1, ½, ⅓, and ¼. Solid $m\phi$ lines are for terminating impedances $z_1 = r_1 = 4$, 3, 2, and 1. Dotted $m\phi$ lines are for terminating impedances $z_1 = r_1 = 1$, ½, ⅓, and ¼.

beam, V_0, can be found. The general equation relating the potential V_0 to the distance x and beam diameter $2s$ will now be derived.

The characteristic beam impedance, W, and phase constant, β, for any transformer may be written[6] as

$$W(\phi) = W = A \frac{p}{s} V_0^{3/4}, \tag{33}$$

$$\frac{d\phi}{dz} = \beta = B \frac{p}{s} V_0^{-3/4} \tag{34}$$

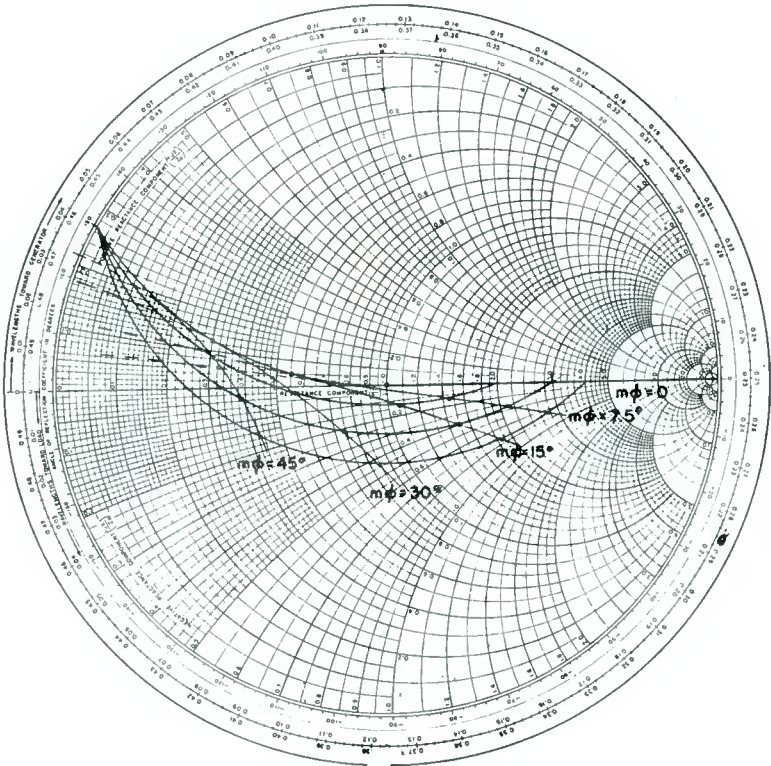where $\qquad A = \dfrac{1.042 \times 10^9}{f I_0^{1/2}},$

and $\qquad B = 5.49\, I_0^{1/2}.$

Fig. 7—Normalized impedance $z = Z/W$ of an exponential transformer with characteristic impedance $W = W_1 e^{k\phi}$ for $k = 0.4$, ($\frac{1}{4} \leq r_1 \leq 4$). Solid $m\phi$ lines are for terminating impedances $1 \leq (z_1 = r_1) \leq 4$. Dotted $m\phi$ lines are for terminating impedances $\frac{1}{4} \leq (z_1 = r_1) \leq 1$.

I_0 is in milliamperes, V_0 in volts and f in cycles per second. From Equations (33) and (34), both ϕ and β are seen to be functions of V_0, s, and p. The plasma frequency reduction factor, p, is in turn a function of V_0, s, and p.

From Equation (33), ϕ may be expressed as a function of W which is in turn a function of V_0, s, and p;

$$\phi = \phi(V_0, s, p). \tag{35}$$

From Equations (34) and (35),

$$\beta = B \frac{p}{s} V_0^{-3/4} = \frac{d\phi}{dx} = \left[\frac{\partial\phi}{\partial p} \frac{\partial p}{\partial V_0} + \frac{\partial\phi}{\partial V_0} + \frac{\partial\phi}{\partial s} \frac{\partial s}{\partial V_0} \right] \frac{dV_0}{dx}, \tag{36}$$
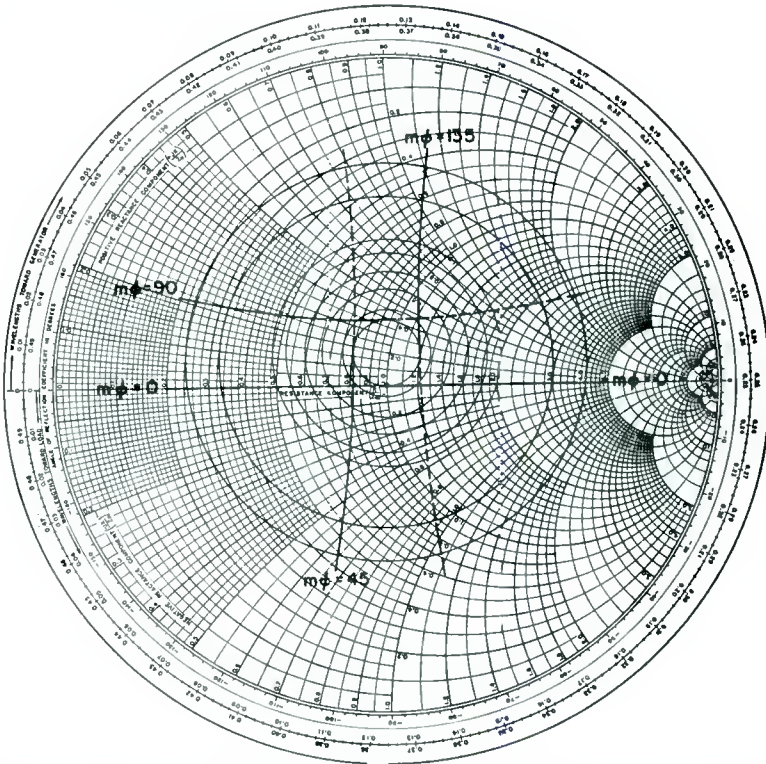

Fig. 8—Normalized impedance $z = Z/W$ of an exponential transformer with characteristic impedance $W = W_1 e^{k\phi}$ for $k = 1$, $(\frac{1}{4} \leqq r_1 \leqq 4)$. Solid $m\phi$ lines are for terminating impedances $1 \leqq (z_1 = r_1) \leqq 4$. Dotted $m\phi$ lines are for terminating impedances $\frac{1}{4} \leqq (z_1 = r_1) \leqq 1$.

and therefore

$$Bx = \int_{V_{0_1}}^{V_0} \frac{s}{p} V_0^{3/4} \left[\frac{\partial\phi}{\partial p} \frac{\partial p}{\partial V_0} + \frac{\partial\phi}{\partial V_0} + \frac{\partial\phi}{\partial s} \frac{\partial s}{\partial V_0} \right] dV_0. \tag{37}$$
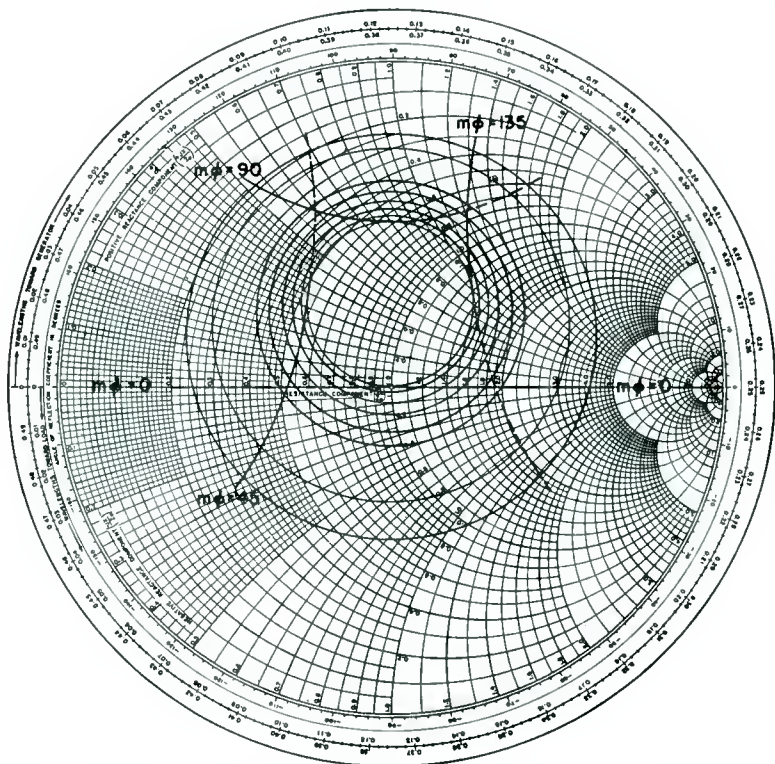
The expression in brackets in Equation (37) is a function of V_0 and s only, according to Equation (35), since p is a function of V_0 and s. In Equation (37), therefore, the potential V_0 and beam diameter, $2s$, along the beam is related to the distance x along the beam. For the case of a strong confining magnet field, $s = $ constant and Equation (37) reduces to

Fig. 9—Normalized impedance $z = Z/W$ along an exponential transformer with characteristic impedance $W = W_1 e^{k\phi}$ for values of $k = 0.6$ to 4 and for perfect initial match, $r_1 = 1$.

$$Bx = \int_{V_{0_1}}^{V_0} \frac{V_0^{3/4}}{p} s \left[\frac{\partial \phi}{\partial p} \frac{\phi p}{\phi V_0} + \frac{\partial \phi}{\partial V_0} \right] dV_0. \qquad (38)$$

This equation will now be solved for the exponential transformer.

Potential Distribution in an Exponential Transformer

For this transformer, the increase of impedance with phase angle ϕ is exponential according to definition, Equation (20),

$$W = W_1 e^{k\phi}. \tag{20}$$

From Equations (20) and (33),

$$\phi = \frac{1}{k} \ln \frac{W}{W_1} = \frac{1}{k} \ln \frac{\dfrac{p}{s} V_0^{3/4}}{\dfrac{p_1}{s_1} V_{0.}^{3/4}}. \tag{39}$$

And, since p_1, s_1, and V_{0_1} (i.e., magnitudes of p, s, and V_0 at the "cathode plane") are constant, Equation (39) can be written

$$\phi = \frac{1}{k} \ln \frac{\left(\dfrac{p}{s}\right) V_0^{3/4}}{C}. \tag{40}$$

Assuming strong focusing, $\partial s/\partial V_0 = 0$ and from Equations (38) and (40),

$$Bx = \int_{V_{0_1}}^{V_0} \frac{s}{kp^2} \left[V_0^{3/4} \frac{\partial p}{\partial V_0} + p \frac{3}{4} V_0^{-1/4} \right] dV_0. \tag{41}$$

For the case of a relatively thick beam, the plasma frequency reduction factor, p, is unity and $\partial p/\partial V_0 = 0$. The solution to Equation (41) in this case becomes

$$kB \frac{x}{s} = V_0^{3/4} - V_{0_1}^{3/4}. \tag{42}$$

In reality, for a finite diameter beam, $p \leq 1$; the relationship between p and V_0 must be known in order to solve Equation (41). An approximate explicit expression for $p(V_0)$ has been proposed by S. Bloom* for the fundamental mode. It is accurate within a few per cent in the range of interest, and will be used to solve Equation (41). The derivation of the potential distribution in the exponential transformer from Equation (41) is given in Appendix I and leads to

---

* Private communication.

$$[I_0]^{1/2} k \frac{x}{s} = \frac{1}{5.49h_1} \left\{ \frac{(1 + h_1 V_0^{3/4})^{3/2} - (1 + h_1 V_{0_1}^{3/4})^{3/2}}{3} \right.$$

$$\left. + (1 + h_1 V_0^{3/4})^{1/2} - (1 + h_1 V_{0_1}^{3/4})^{1/2} \right\}$$

$$(43)$$

where x and s are in meters, I_0 is in milliamperes, and
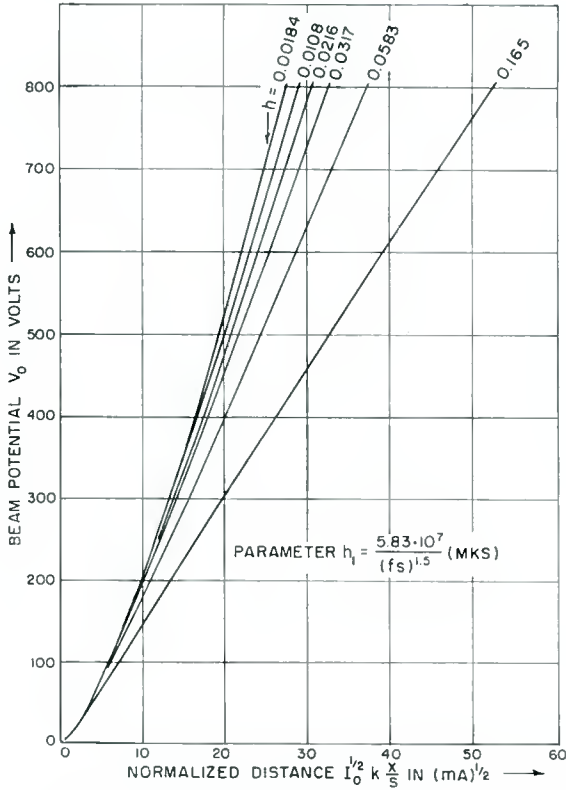

Fig. 10—Potential variation along a beam in an exponential gun for $V_{0_1} = 0.1$ volt.

$$h_1 = \frac{5.83 \times 10^7}{(fs)^{3/2}}. \qquad \text{(mks)} \qquad (44)$$

Figure 10 is a plot of V_0 versus normalized axial distance $\times I_0^{1/2} k/s$ with h_1 as a parameter and with initial d-c beam potential $V_{0_1} = 0.1$ volt.

Low-Noise Gun Requirements

A "low-noise gun," defined in the introduction as the transformer region (2) of an electron gun, should satisfy the following requirements:

(a)   It should act as a matching transformer between those impedances, at the cathode and circuit input planes, which minimize the noise figure of the fundamental wave;

(b)   It should match not only the fundamental noise wave but also higher-order noise wave modes;

(c)   The matching requirements (a) and (b) should be accomplished not only at a single operating frequency but preferably over the whole frequency range of the amplifier;

(d)   It should not add any noise to the beam noise which enters the transformer region (2);

(e)   It should satisfy conditions (a), (b), (c), and (d) independent of amplifier gain (i.e., independent of beam current).

For the following discussion, it shall be assumed that the optimum boundary beam impedances are those given by Bloom and Peter.[2] The normalized cathode impedance for the fundamental noise wave is

$$z_1 = r_1 = 1.52 \times 10^{-9} f \frac{T_c^{1/4}}{J_{0_1}^{1/2}}, \tag{45}$$

where T_c is the cathode temperature in degrees Kelvin and J_{0_1} is the cathode current density in amperes per square meter.

The normalized impedance at the circuit input plane z yielding minimum noise for the fundamental wave is given in Figure 16 on the $n = 1$ curve. The matching of higher-order modes is discussed later.

From theoretical and experimental evidence, the following conditions were found to serve as a useful guide toward achieving the above-stated requirements (c), (d), and (e):

(1)   The beam should be accelerated in a smooth fashion avoiding potential discontinuities;

(2)   The termination at the cathode end of the transformer should be chosen close to perfect match;

(3)   The gun measured in plasma wavelengths should be short;

(4)   The noise-current (or voltage) standing-wave ratio throughout the transformer should not exceed either the value required at the circuit input plane or the standing-wave ratio caused by mismatch at the cathode, whichever is larger.

The Exponential Gun

Of the various transformers described above only the exponential gun can satisfy the desired conditions for a low-noise gun.

The diode transformer by itself (Figures 2 and 3) cannot match the required impedance z (Figure 16, $n = 1$) at the circuit input plane.[22] In order to obtain a match, cascaded velocity jumps and drift regions can be employed. The result is the velocity-jump gun described by Watkins.[23] A procedure for designing velocity jump guns has been given by Buchmiller et al.[22]

Comparing the velocity-jump gun with the desired low-noise gun, we find two major shortcomings. Due to the inherent presence of abrupt potential changes, there exist lens effects which have been shown to increase the beam noisiness.[24] Also, once the lengths of the drift tubes are chosen, an optimum match is obtained for one particular beam current, and a limited frequency band only.

The "exponential transformer," characterized by a single exponent k, on the other hand, does not suffer from these disadvantages. As may be seen from Figure 10, the potential rise with distance for such a "gun" is almost linear. Lens effects therefore are minimized. Also, the exponential transformer is known for its broad-band match-. ing qualities.

It is difficult in practice to realize a true exponential gun. Close to the cathode, in the vicinity of the potential minimum (if there is one), the potential distribution corresponds to that of a space-charge-limited diode. Even if it were technically feasible to produce an immediately rising characteristic beam impedance, it would be of questionable value. It appears that an initially constant impedance is preferable, keeping in mind the stated requirements (recent studies[7-10] indicate that a constant impedance drift space in the "cathode region" can have a major noise-reducing effect). If this reasoning is correct, nature, for a change, makes the realization of the better solution easier.

In Figure 11 the normalized characteristic beam impedance for the four basic transformers is plotted as a function of the phase angle, ϕ. The drift region, the "Bessel Line" transformer, the space-charge-limited diode, and the exponential transformer impedance variation are shown for conditions of perfect initial match. This is

[22] L. D. Buchmiller, R. W. DeGrasse and G. Wade, "Design and Calculation Procedures for Low-Noise Traveling-Wave Tubes," *Trans. I.R.E. PGED*, p. 234, July, 1957.

[23] D. A. Watkins, "Noise Reduction in Beam Type Amplifiers," *Proc. I.R.E.*, Vol. 40, p. 65, January, 1952.

[24] R. C. Knechtli, "Effect of Electron Lenses in Beam Noise," *Trans. I.R.E. PGED*, p. 84, April, 1958.

a plot of Equations (6), (10), (16) and (20). Considering the impedance of the infinite plane parallel diode (Figure 11), a smooth initial increase of impedance is seen to exist. The phase derivative of this impedance is also increasing smoothly in the diode. This diode condition is always present in the cathode region for parallel electron flow. The impedance variation in a practical exponential gun, there-

Fig. 11—Normalized characteristic beam impedance W/W_1 for various space-charge-wave transformers, plotted versus phase angle, ϕ, in plasma-frequency radians.

fore, will correspond to space-charge-limited flow near the cathode. It will go smoothly into an exponential impedance variation for the rest of the gun. This gun is a modified exponential gun referred to here simply as the "exponential gun." The characteristic beam impedance variation for this gun is also shown in Figure 11.

The potential distribution $V_0(z)$ for the exponential gun is initially

identical with that of parallel space-charge-limited electron flow. This is clearly necessary if the same beam-current density is postulated in both cases. A short distance away from the cathode, however, the potential in the exponential impedance gun increases in an approximately linear fashion (Figure 10). The electron beam in the exponential gun, therefore, has a tendency to expand since the potential does not increase sufficiently to keep it parallel. Magnetic confinement is required to hold the beam diameter constant.

The Three-Region Gun—A Practical Low-Noise Gun

The three-region gun was designed with the idea of obtaining as flexible a transformer as possible. A series of three or more accelerating electrodes allows greatly varied potential distributions along

Fig. 12—Equipotential surfaces and beam profile in a three-region low-noise gun.

the beam axis (see Figure 12). The cathode-to-first-anode region was designed as a parallel-flow Pierce-type gun. However, in order to test convergent and divergent initial electron flow also, a design with protruding cathode was chosen; this allows considerable flexibility in the initial electron-flow conditions. Recently, more detailed studies indicated that positioning of an additional electrode near the cathode can produce an electron beam whose initial divergence angle can be adjusted over a large range (typically, deviations of $\pm$ 10° can be obtained).

As a result of extensive experimental studies, it was found that an optimum noise factor in a typical 3,000-megacycle traveling-wave tube was consistently obtained with conditions closely approaching those of Figure 12. A detailed analysis was made of the impedance transformation existing along the beam in this gun. The potential, $V_0(x)$, along the axis was obtained from electrolytic tank measure-

ments (Figure 13). The beam diameter, $2s$, was then computed under
the assumption of laminar flow in a confining uniform magnetic field
of 500 gausses. From this, the phase constant, β, was determined
(from Equation (34)) as a function of the beam voltage, V_0, and
beam diameter, $2s$, (Figure 14). Finally, the characteristic beam
impedance W, Equation (33), was computed. A graphical integration
of β yields ϕ as a function of distance and voltage. Since the char-
acteristic beam impedance *and* the phase angle are known as functions
of the voltage, the characteristic beam impedance may be plotted
versus the phase angle. Figure 15 is a plot of the impedance W as

Fig. 13—Potential variation along beam in low-noise gun of Figure 12.

a function of phase angle ϕ between a plane .015 inch in front of the
cathode and the third anode plane.

With this information about the three-region gun, the transfer
matrix of Equation (1) was computed in three ways:

(1) by approximating the actual impedance function $W(\phi)$ by
a large number of short drift sections and small impedance
jumps,[22] Equation (9) ;

(2) by approximating $W(\phi)$ by a small number of exponential
transformers with different exponents k, Equations (20),
(24), and (25) ;

(3) by approximating the plotted impedance function log W

$$\beta / f I_0^{\frac{1}{2}} = 5.815 \cdot 10^4 \left(\frac{p}{\gamma_0 s}\right) V_0^{-\frac{5}{4}}$$

Fig. 14—Normalized phase parameter $\beta/fI_0^{\frac{1}{2}}$ along low-noise gun of Figure 12, where β is in seconds per meter, f is in cycles per second, and I_0 is in milliamperes.

Fig. 15—Normalized characteristic beam impedance for low-noise gun of Figure 12.

versus ϕ by a single exponential transformer (average straight line on this plot).

The interesting result of this study was the discovery that the "crude" approximation No. (3) gave approximately the same matrix as the more exact solutions, Nos. (1) and (2). It was concluded that the three-region gun must represent an approximation to an exponential gun of the modified type (at least between first and third anode).
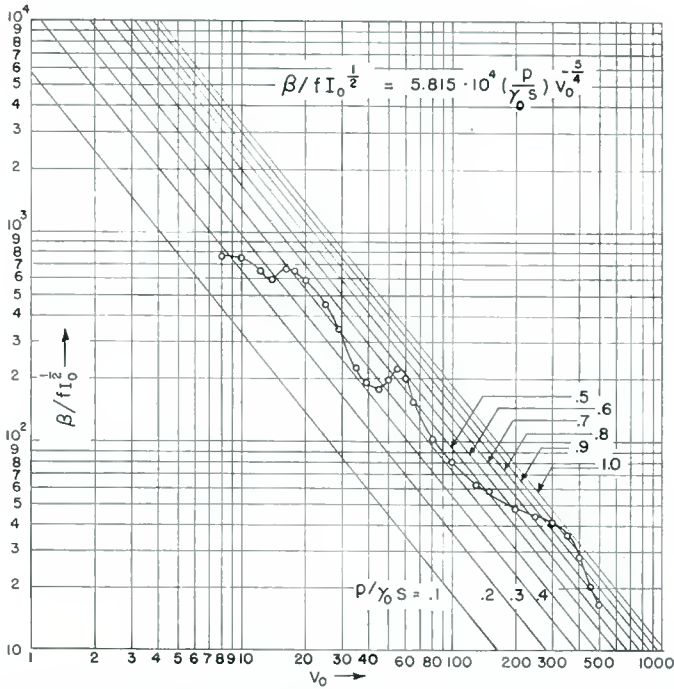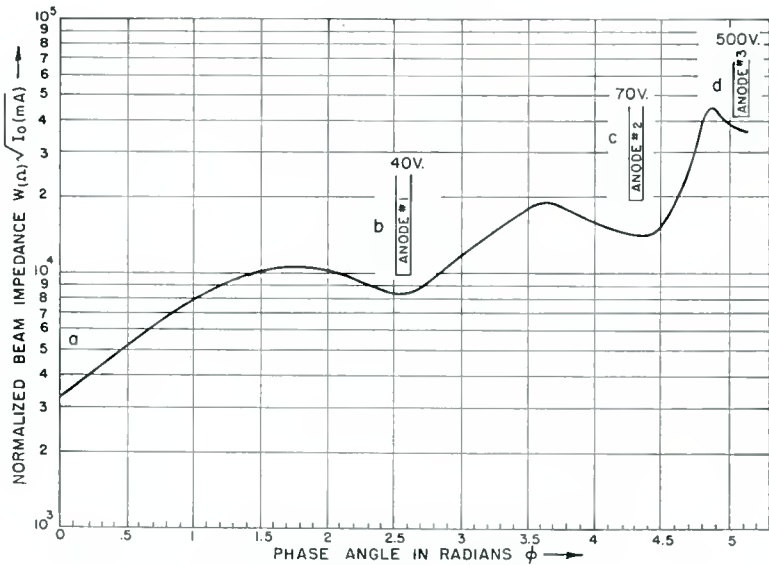
Since the space-charge in the beam is relatively small, the potential distribution along the beam is very nearly the one produced by the electrode potentials, with the exception of the cathode region. In the interelectrode space the potential rise is essentially linear, and the field gradient is constant. Hence, each interelectrode section approximates an exponential transformer section with some value of k.

A three-region gun (Figure 12) therefore consists of three exponential transformer sections a-b, b-c, and c-d, the first of which is of the modified type. One suspects that with an initial match at the cathode end, the best match at the circuit input plane would be obtained if, approximately,

$$k_{ab} = k_{bc} = k_{cd}, \tag{46}$$

i.e., if the field gradient in all three regions is the same. This was borne out by experiments of Knechtli and Beam.[25] The condition for constant gradient throughout the gun required the following electrode voltages:

$$V_{a_1} = 50 \text{ volts}$$
$$V_{a_2} = 270 \text{ volts}$$
$$V_{a_3} = 650 \text{ volts.}$$

For these potentials the measured impedances were indeed close to perfect match. Departure from these constant gradient conditions lead to a higher standing wave ratio.

These experiments indicate that an increase of potential on any electrode shortens the phase angle of the gun. Furthermore, a change in first anode potential, V_{a_1}, has a larger effect on the circuit input impedance, z, than an equal change in second anode potential, V_{a_2}.

The reasons for these effects become apparent in the following section.

---

[25] R. C. Knechtli and W. R. Beam, "Performance and Design of Low Noise Guns for Traveling-Wave Tubes," *RCA Review*, Vol. XVII, p. 410, September, 1956.

Higher-Order Space-Charge-Wave Modes

As discussed briefly in the Introduction, a multiple infinity of higher-order modes is excited by the fluctuation of emission at the transformer-region input. (Most traveling-wave-tube low-noise amplifiers use axially symmetric circuits interacting with an axial beam. In this case, only modes with axial symmetry need to be considered.)

For each higher-order mode, there exists an optimum boundary condition at the circuit input, just as for the fundamental mode. It is assumed in this paper that this, as well as the condition at the transformer input, is known for each mode.

As mentioned earlier, the optimum boundary conditions at both end of the transformer region are not yet definitely established. However, good approximations have been given.[5,26]

In order to demonstrate the feasibility of multimode matching with an exponential gun, the following simplifying assumptions are made:

(1) Beam's[26] solution for higher-order-mode matching is used as boundary conditions at the circuit input plane;

(2) A perfect match is assumed for all modes at the cathode plane;

(3) A single exponential gun is assumed to be used as transformer between cathode and circuit input plane.

It should be understood, however, that the same basic procedure will still hold if the boundary conditions are somewhat different from the ones assumed in the following example. Similarly, the proposed method is still applicable if a different type of transformer (e.g., a 3-region gun) is used. R. C. Knechtli[27] measured the fundamental and first higher-order standing-noise wave in a beam which was produced by a 3-region gun. His measurement showed that, at least, these two modes could be optimized if the potential distribution in the gun were properly chosen.

Assumed Boundary Conditions

The optimum normalized impedances, z_n, at the circuit input plane are given in Figure 16. These curves give the minimum noise factor for the various modes as computed by Bloom[5] for the funda-

---

[26] W. R. Beam, "Extension of the Effect of Initial Noise Current and Velocity Correlation on the Noise Figure of Traveling-Wave Tubes," *RCA Review*, Vol. XVI, p. 458, September, 1955.

[27] Research and Development on Microwave Generators, Mixing Devices and Amplifiers, Signal Corps Contract No. DA36-039-so-64443, Quarterly Report No. 6, April, 1956.

mental $n = 1$, and Beam[26] for the higher-order modes $n = 2$, 3, and 4. The impedances z_n are given for different values of Bloom's[5] parameter $\overline{y}$ which is a function of Pierce's[28] space-charge parameter, QC, and circuit-loss parameter, d. For a chosen $\overline{y}$, for example, the optimum impedance z_1 for lowest noise induced by the fundamental space-charge wave is found on the $n = 1$ curve of Figure 16. Correspondingly, optimum impedance values z_2, z_3, for the higher-order modes lie on the $n = 2$, $n = 3$, etc., curves. The dotted lines of Figure 16 connect the optimum normalized impedance z_1, z_2, z_3 of the various modes for a number of parameter values, $\overline{y}$.

Fig. 16—Optimum normalized beam impedances z_n at the circuit input plane for minimum noise in the fundamental ($n = 1$) and higher-order space-charge-waves ($n = 2$, 3, 4) after Bloom[5] and Beam.[26]

At the cathode plane a perfect match is assumed for each mode.

An exponential transformer is used for matching the impedances at the two boundaries. In Figure 17, the matched exponential transformer characteristics for $k < 2$ of Figure 9 have been superimposed on the curves of Figure 16. From Figure 17, the impedances leading to minimum noise figure for each mode are obtained in terms of the transformer parameters, k and ϕ_0. For the fundamental mode im-

---

[28] J. R. Pierce, *Traveling-Wave Tubes*, D. Van Nostrand Co., Princeton, N. J., 1950.

pedance z_1, the required k_1 and ϕ_{0_1} are found and similarly k_n and ϕ_{0_n} are found for each higher-order mode impedance, z_n. The total angle, ϕ_n, is determined by the number of times the impedance circle z_n is traversed, each time adding π to the angle ϕ_{0_n}; therefore,

$$\phi_n = q_n\pi + \phi_{0_n}, \tag{47}$$

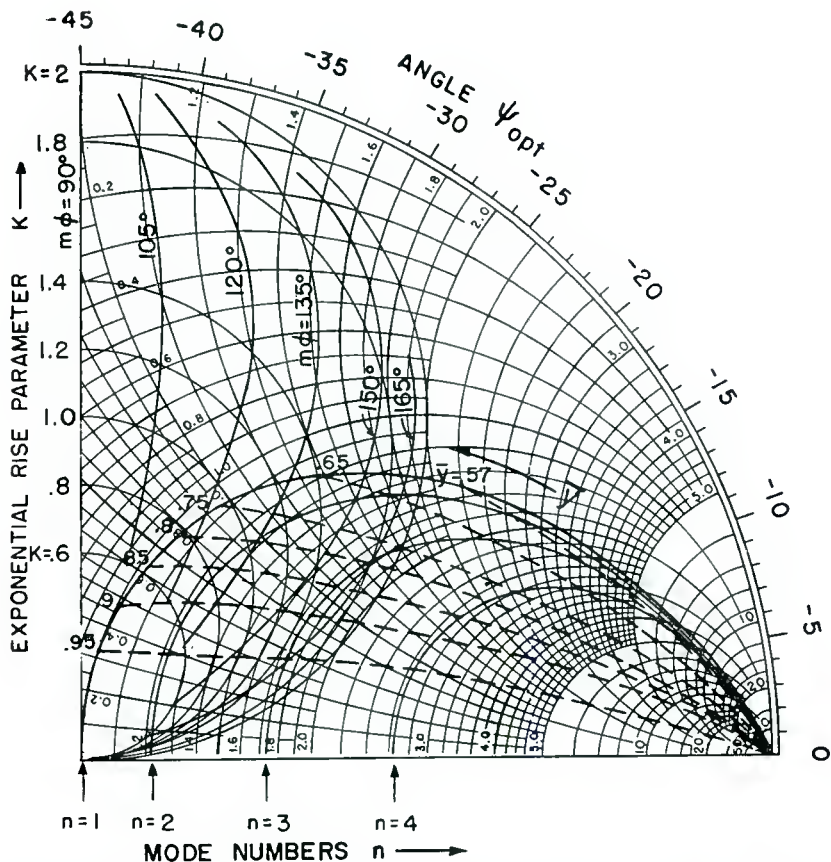Fig. 17—A plot of normalized impedance curves for an exponential beam transformer ($k < 2$ and perfect initial match, see Figure 9) on the curves of optimum beam impedance according to Figure 16.

where q_n is a positive integer. Note that the required relationships for matching of the various modes are obtained from Figure 17:

$$k_1 < k_2 < k_3 < \cdots < k_n, \tag{48}$$

$$\phi_{0_1} < \phi_{0_2} < \phi_{0_3} < \cdots < \phi_{0_n}. \tag{49}$$

Multimode Wave Propagation in the Exponential Transformer

The characteristic beam impedance for the nth mode in an exponential gun as obtained from Equations (20) and (33) is

$$W_n = W_{n_1} e^{k_n \phi_n} = A \frac{p_{n_1}}{s_1} V_{0_1}^{3/4} \cdot e^{k_n \phi_n} = A \frac{p_n}{s} V_0^{3/4}. \tag{50}$$

At the cathode plane 1, the following is assumed:

$$p_{1_1} = p_{2_1} = p_{3_1} = p_{4_1} = \cdots p_{n_1} = 1. \tag{51}$$

Assuming confined flow (beam diameter $2s = \mathrm{const.}$), from Equation (50),

$$k_n \phi_n = \ln \left[p_n \left(\frac{V_0}{V_{0_1}} \right)^{3/4} \right]. \tag{52}^*$$

A plot of the plasma-frequency reduction factor, p, for the various modes (Figure 18) shows that for any given $\gamma_0 s$

$$p_1 > p_2 > p_3 > \cdots > p_n, \tag{53}$$

and therefore with Equation (52)

$$k_1 \phi_1 > k_2 \phi_2 > k_3 \phi_3 > \cdots > k_n \phi_n. \tag{54}$$

It will now be shown that for an exponential gun,

$$k_1 < k_2 < k_3 < \cdots < k_n. \tag{55}$$

---

* Equation (52) explains the experimental behaviour described earlier, namely that a change in first-anode potential affects the transformation in the gun more than an equal change of potential on any other electrode. In the cathode–first-anode region, V_{0_1} is the potential at initial plane 1 which is small. The ratio of first-anode potential V_{a_1} to V_{0_1} is therefore large. Since this ratio is large, $k_n \phi_n$ is also large as may be seen from Equation (52). A change in V_{a_1} therefore changes considerably the total angle $k_n \phi_n$ in the gun. In the first-anode–second-anode region the initial plane potential is the first anode potential V_{a_1} and the final potential is second anode potential V_{a_2}. The ratio V_{a_2}/V_{a_1} is therefore small and changes in V_{a_2} have little effect on the total transformation in the gun.

Approximating the plasma frequency reduction factor, p, for mode, n, by

$$p_n = (1 + h_n V_0^{3/4})^{-1/2}$$

yields the following relation between h_n values for different modes (see Appendix II, Equation (73)):

$$h_1 < h_2 < h_3 < \cdots < h_n \qquad (56)$$

Fig. 18—Plasma frequency reduction factor p_n for the lowest four modes ($n = 1$ is fundamental mode; dashed curves are an approximation obtained from an empirical analytic expression for p_n).

for any given value V_0. Applying Equation (56) to an exponential gun where values V_0 and $I_0^{1/2}x/s$ are fixed (Figure 10) shows that Equation (55) is satisfied.

Since, further, Equations (55) and (54) are satisfied simultaneously in the exponential transformer, it follows that

$$\phi_1 > \phi_2 > \phi_3 > \cdots \phi_n > \phi_{n+1}. \qquad (57)$$

Multimode Matching with the Exponential Gun

The required condition of Equation (48) is qualitatively satisfied in an exponential transformer according to Equation (55).

In order to find the conditions to also satisfy Equation (49), Equation (47) is introduced into Equation (57), yielding

$$(\pi q_1 + \phi_{0_1}) > (\pi q_2 + \phi_{0_2}) > (\pi q_3 + \phi_{0_3}) > \cdots \qquad (58)$$

Equations (58) and (49) can only be satisfied simultaneously if

$$q_1 > q_2 > q_3 > \cdots > q_n. \qquad (59)$$

All q_n are positive integers and $\phi_{0_n} < \pi$ (Figure 17).

It will now be shown how Equations (55) and (58) can be satisfied for 2 or 3 modes in an exponential gun. The method used is one of successive approximation.

(a) The first approximation is obtained by arbitrarily choosing a "reasonable" value of Bloom's[5] parameter, $\bar{y}$.

(b) Using Figure 17 in the manner described earlier, the required optimum values k_1, k_2, k_3 and ϕ_{0_1}, ϕ_{0_2}, ϕ_{0_3} are obtained with the chosen value of $\bar{y}$ for the first three modes, $n = 1$, 2, and 3.

(c) Values q_1, q_2, and q_3 are now chosen to satisfy Equation (59). Transit angles ϕ_1, ϕ_2, ϕ_3 are then found, Equation (47). With the values k_n determined in (b), the ratios $k_1\phi_1/k_2\phi_2$ and $k_1\phi_1/k_3\phi_3$ are found.

(d) In Figure 19, the ratio $k_1\phi_1/k_2\phi_2$ is plotted versus V_0, using Equation (52) and the approximation for p (Equation (73), Appendix II). In Figure 20, the analogous ratio $k_1\phi_1/k_3\phi_3$ is plotted. For both plots, a value of $V_{01} = 0.1$ volt at the cathode plane 1 was chosen.

One must now find such a value V_0 in Figures 19 and 20 for which $k_1\phi_1/k_2\phi_2$ and $k_1\phi_1/k_3\phi_3$, as found in (c), intersect h curves with the same parameter value in both figures. If no V_0 value can be found to satisfy this condition, a new $\bar{y}$ value must be chosen and steps (a) to (d) have to be repeated until a satisfactory $\bar{y}$ value is arrived at.

(e) The previous results may be verified with the aid of Figure 21. The h_n values corresponding to the intersections of $k_n\phi_n$ and V_0 are found from this graph. If the successive approximation procedure was successful, the h values for the first 3 modes should be related, according to Appendix II, as follows:

$$h_1 : h_2 : h_3 = 1 : 7.5 : 20.9. \qquad (60)$$

(f) With the values V_0 and h_n determined, $I_0^{1/2} kx/s$ is found

Fig. 19—Ratio $k_1\phi_1/k_2\phi_2$ plotted as a function of V_0 with h as a parameter.

from Figure 10.* The beam current density, J_0, is given by the initial condition of perfect match (Equation (45)). The beam radius, s, is determined by h_1, for a given frequency, f, according to Equation (73), Appendix II. The length, x, of the exponential gun can therefore be computed.

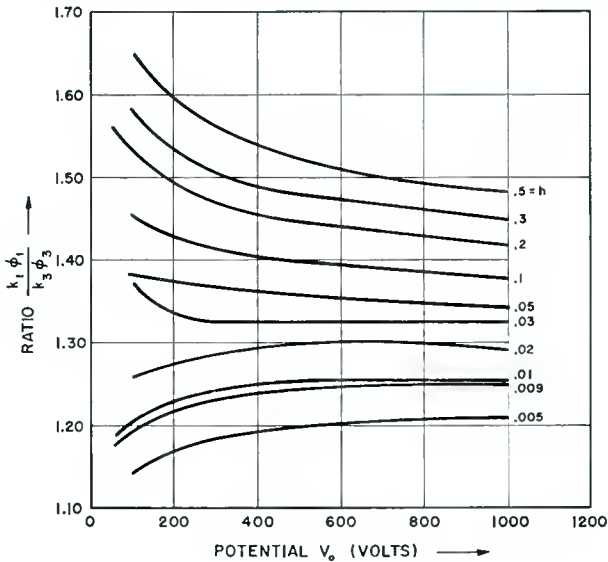This design procedure will be illustrated with a practical example.

Fig. 20—Ratio $k_1\phi_1/k_3\phi_3$ as a function of V_0 with h as a parameter.

* In Figure 10, $k = k_1$. This ensures that the gun is optimized perfectly for the fundamental mode. This is important since the fundamental mode contributes by far the largest portion to the total noise factor.

Example of Low-Noise Gun Design

Problem: Design a low-noise gun for 3,000 megacycles to operate at about 600 volts circuit potential, optimized for the first three modes. Find the gun length, x, the total d-c current, I_0, the beam diameter, $2s$, and the circuit potential, V_0.

Solution: An exponential gun shall be used with boundary conditions as stipulated above. Following the design procedure described earlier, one goes through the following steps:

(a)   A value $\overline{y} = 0.8$ is chosen as a "reasonable" starting point.

Fig. 21—Plot of $k\phi$ as a function of V_o with h as a parameter, for $V_{o_1} = 0.1$ volt.

(b)   From Figure 17

for $n = 1$: $k_1 = .78,$ $\phi_{0_1} = 115°,$

$n = 2$: $k_2 = 1.05,$ $\phi_{0_2} = 135°,$

$n = 3$: $k_3 = 1.8,$ $\phi_{0_3} = 165°.$

(c)   Choosing

$$q_1 = 2, \ q_2 = 1, \ q_3 = 0$$

in agreement with Equation (59),

$$\phi_1 = 360° + 115° = 475°,$$
$$\phi_2 = 180° + 135° = 315°,$$
$$\phi_3 = 165°,$$

and

$$\frac{k_1\phi_1}{k_2\phi_2} = 1.12, \quad \frac{k_1\phi_1}{k_3\phi_3} = 1.25.$$

(d)   These ratios are entered in Figures 19 and 20. At no value of V_0 do both these ratios intersect with a curve of the same parameter h. Steps (a) to (d) are therefore repeated with a new value of $\overline{y}$.

(a)   As a new value, slightly changed, $\overline{y} = .81$ is tried.

(b)   From Figure 17,

$$n = 1 : \qquad k_1 = .752, \qquad \phi_{0_1} = 114°,$$
$$n = 2 : \qquad k_2 = .99, \qquad \phi_{0_2} = 133°,$$
$$n = 3 : \qquad k_3 = 1.75, \qquad \phi_{0_3} = 163°.$$

(c)   Using the same values, $q_1 = 2$, $q_2 = 1$, $q_3 = 0$, as in the first attempt leads to

$$\phi_1 = 360 + 114° = 474°,$$
$$\phi_2 = 180 + 133° = 313°,$$
$$\phi_3 = 0 + 163° = 163°,$$

and

$$\frac{k_1\phi_1}{k_2\phi_2} = 1.152, \qquad\qquad \frac{k_1\phi_1}{k_3\phi_3} = 1.252.$$

(d)   These ratios are now again entered in Figures 19 and 20. It is seen that for a circuit voltage of 600 volts, these ratios intersect quite closely the same h curves, $h = .009$, in both figures. Further approximation does not seem necessary. A further check is next made to assure that the solution is adequate before proceeding to the gun design.

(e)   The values

$$k_1\phi_1 = 6.2, \qquad k_2\phi_2 = 5.4, \qquad k_3\phi_3 = 4.95$$

are entered on Figure 21. The intersection points at $V_0 = 600$ volts yield the h_n values for the three modes:

$$h_1 = .009, \qquad h_2 = .067, \qquad h_3 = .19.$$

It is found that these values are very closely related as $h_1 : h_2 : h_3 = 1 : 7.5 : 20.9$, as required according to Equation (60).

(f)   With $V_0 = 600$ volts and $h_1 = .009$, $I_0 \frac{1}{2} kx/s = 22.87$ is found from Figure 10. The condition of initial match at the cathode ($z_1 = 1$ in Equation (45)) requires a beam-current density of $J_{0_1} = .065$ ampere per square meter for a cathode temperature of $1,000°$K. The beam

radius $s = 1.15$ millimeters is obtained from Equation (73), using $f = 3,000$ megacycles. We may therefore find the current $I_0 = J_0 s^2 \pi = 2.76$ milliamperes. Finally, the gun length, x, can now be determined. From Figure 10, $I_0^{1/2} k_1 x/s = 22.87$; $x = 2.1$ centimeters.

APPENDIX I—COMPUTATION OF POTENTIAL DISTRIBUTION IN AN EXPONENTIAL GUN

In the case $p \leq 1$, which normally exists in the gun region, the relationship for the potential distribution may be obtained from Equation (41).

$$Bx = \frac{3}{4} \int_{V_{0_1}}^{V_0} s \frac{V_0^{-1/4}}{pk} dV_0 + \int_{V_{0_1}}^{V_0} \frac{sV_0^{3/4}}{p^2 k} \frac{\partial p}{\partial V_0} dV_0. \qquad (61)$$

An analytical relationship between p and V_0 must be used in order to solve Equation (61) explicitly. An approximate expression for the dependence of p on V_0 for the fundamental mode may be written:[*]

$$p_1 = \frac{1}{[1 + 2/(\gamma_0 s)^{3/2}]^{1/2}}, \qquad (62)$$

where

$$\gamma_0 = \frac{w}{v_0} = \frac{2\pi f}{\sqrt{2 \frac{e}{m} V_0}}. \qquad (63)$$

Substitution of Equation (63) into (62) yields

$$p_1 = \frac{1}{(1 + h_1 V_0^{3/4})^{1/2}}, \qquad (64)$$

where

$$h_1 = \frac{5.83 \times 10^7}{(fs)^{3/2}} \text{ (mks)}. \qquad (65)$$

Equation (62) is plotted in Figure 18 as p_1 versus $\gamma_0 s$ and compared to the accurate value of p_1. The agreement is seen to be excellent. In practical cases the beam radius oscillates around the equilibrium value and neither s nor p_1 is known exactly. The approximation for p is therefore very good for most practical cases.

---

[*] Similar to Bloom's approximation.

In the first integral on the right-hand side of Equation (61), p_1 may be replaced by the value given in Equation (64). In the second integral on the right-hand side of Equation (61) it is more convenient to have p as the variable. At the limits V_{0_1} and V_0, p has the values p_{1_1} and p_1 respectively. Since p is a function of V_0 and s

$$dp = \frac{\partial p}{\partial V_0}\, dV_0 + \frac{\partial p}{\partial s}\, ds \tag{66}$$

for strong confinement, the second term on the right-hand side of Equation (66) drops out and

$$dp = \frac{\partial p}{\partial V_0}\, dV_0. \tag{67}$$

From Equation (64) $V_0{}^{3/4}$ may be expressed in terms of p_1:

$$V_0{}^{3/4} = \frac{1 - p_1{}^2}{p_1{}^2 h_1}. \tag{68}$$

By use of Equations (64), (67), and (68), Equation (61) may be written

$$Bx\,\frac{k}{s} = \frac{3}{4} \int_{V_{0_1}}^{V_0} V_0{}^{-1/4}\, [1 + h_1 V_0{}^{3/4}]^{1/2}\, dV_0 + \int_{p_{1_1}}^{p_1} \frac{1 - p^2}{h_1 p^4}\, dp. \tag{69}$$

The first integral of Equation (69) may be evaluated

$$\frac{2}{3h_1}\, [1 + h_1 V_0{}^{3/4}]^{3/2} \Bigg|_{V_{0_1}}^{V_0} = \frac{2}{3h_1}\, \{[1 + h_1 V_0{}^{3/4}]^{3/2} - [1 + h_1 V_{0_1}{}^{3/4}]^{3/2}\}. \tag{70}$$

The second integral of Equation (69) yields

$$-\frac{1}{3h_1}\, [(1 + h_1 V_0{}^{3/4})^{3/2} - (1 + h_1 V_{0_1})^{3/4})^{3/2}]$$

$$+\frac{1}{h_1}\, [(1 + h_1 V_0{}^{3/4})^{1/2} - (1 + h_1 V_{0_1}{}^{3/4})^{1/2}]. \tag{71}$$

Substituting the value of $B = 5.49 I_0^{1/2}$ (where I_0 is in milliamperes) and Equations (70) and (71) into Equation (69) yields a relation between V_0 and normalized distance;

$$\frac{kx}{s} I_0^{1/2} = \frac{1}{16.47 h_1} \{ (1 + h_1 V_0^{3/4})^{3/2} - (1 + h_1 V_{0_1}^{3/4})^{3/2} \}$$

$$+ \frac{1}{5.49 h_1} \{ (1 + h_1 V_0^{3/4})^{1/2} - (1 + h_1 V_{0_1}^{3/4})^{1/2} \}. \tag{72}$$

The dependence of V_0 on $I_0^{1/2} kx/s$ has been plotted in Figure 11 with h as a parameter.

Appendix II

Approximate analytical expressions relating the plasma-frequency reduction factor, p, to V_0 for the higher-order modes hold over a range of $1 < \gamma_0 s < 2$. The expression for the first four modes may be written

$$p_n = \frac{1}{(1 + h_n V_0^{3/4})^{1/2}} \quad (n = 1, 2, 3, 4) \tag{73}$$

where

$$h_1 = 5.87 \times 10^7 \ (fs)^{-3/2}, \cdots$$
$$h_2 = 4.37 \times 10^8 \ (fs)^{-3/2},$$
$$h_3 = 1.22 \times 10^9 \ (fs)^{-3/2},$$
$$h_4 = 2.48 \times 10^9 \ (fs)^{-3/2}.$$

The approximate p values given in Equation (73) are plotted in Figure 18 and compared to the exact p values.

GAIN–BANDWIDTH PRODUCT FOR PHOTOCONDUCTORS*

By

A. ROSE AND M. A. LAMPERT

RCA Laboratories,
Princeton, N. J.

Summary—The gain–bandwidth product for photoconductors is shown quite generally to have the form

$$G \frac{1}{\tau_0} = \frac{1}{\tau_{rel}} M,$$

where G = photoconductive gain, τ_0 = response time, τ_{rel} = dielectric relaxation time, and M is a factor dependent on the trap distribution. For many materials the maximum value of M is shown to be unity. For these materials, the gain–bandwidth product approaches zero as the conductivity of the material approaches zero. The current performance of the Vidicon, Electrofax, Xerography, and light amplifiers is quantitatively matched by $M \cong 1$. So also, most of the data on the simple photocells using CdS and CdSe are consistent with $M \cong 1$. Both analytical and experimental[10,11] evidence indicates that M can be much greater than unity when the density of recombination states near the Fermi level exceeds the density of trapping states. Space-charge current flow in solids plays a key role in the present analysis.

INTRODUCTION

EXTENSIVE experience with relatively insulating photoconductors, both as simple photodetectors and as photoelements in more complex devices, has revealed departures from ideal performance frequently as large as 10^6 fold. These departures are reviewed, and the performance of photoconductors generally is analyzed to show first why these large departures are to be expected for most materials and second what special conditions are necessary to minimize the departures. Finally, the performance of photoconductors in several well known devices is discussed.

STATEMENT OF THE PROBLEM

The sensitivity of a photoconductor is given,[1] quite generally, by the relation†

---

* Manuscript received January 6, 1959.

[1] A. Rose, "Performance of Photoconductors," *Proc. I.R.E.*, Vol. 43, p. 1850, December, 1955.

† The only restriction is that the electric field be less than that required for avalanche breakdown.

$$G = \frac{\tau}{T}, \tag{1}$$

where G is the gain or number of electrons passed through the photo-conductor per absorbed photon, τ is the lifetime of a free electron* (the time it spends in the free states but not the time it spends in traps), and T is the transit time of a free electron from cathode to anode. In the absence of significant trapping effects, the lifetime of a free electron should also be the response time of the photoconductor to changes in light intensity. Equation (1) can then be written:

$$\text{Gain} \times \text{Bandwidth} \equiv G \frac{1}{\tau} = \frac{1}{T}. \tag{2}$$

Ideal performance, as used in this paper, is defined by Equation (2).

Experimentally, Equation (2) is observed to be valid under the conditions that trapping is not significant. This means, usually, that Equation (2) is valid for photoconductors operated in the semiconducting range of resistivities of 10^4 ohm-centimeters or less. The low resistivity may be either a thermal equilibrium value in the dark or a result of ambient illumination. A common example is a CdS photoconductor operated at room illumination or above. The lifetime and response time are equal and in the range of milliseconds.

If, now, the CdS photoconductor has a sufficiently small dark current to be operated in the range of 10^{10} ohm-centimeters (resistivity) or higher, the usual observation is that the response time to changes in light intensity has increased from some milliseconds to some minutes, while the lifetime may be essentially unchanged. Equation (2) is no longer satisfied. The gain–bandwidth product has dropped to 10^{-5} of its value at high lights. The performance is only 10^{-5} of ideal performance.

It is well known that the low performance at low currents is a result of trapping effects such that a change in free carrier concentration must be accompanied by a change in trapped carrier concentration. In this way, the speed of response is reduced. Since, at low currents, the trapped carriers may exceed the free carriers by many powers of ten, the large departures from ideal performance are readily accounted for. It is tempting at this point to say that the removal of traps will insure the same gain–bandwidth product at low lights as is normally observed at high lights. The statement is only partially

---

* In most sensitive photoconductors the current is carried predominantly by one sign of carrier. For definiteness the electron is selected here.

correct. Indeed, the major part of this paper is devoted to showing that for most trap distributions the maximum gain–bandwidth product is insensitive to the density of traps and is proportional mainly to the conductivity. A second part is devoted to the special trap distributions needed to maintain the low-light (or low conductivity) gain–bandwidth product equal to that at highlights (or high conductivities).

The demonstration that the gain–bandwidth product depends in first order on the conductivity was reported independently by Rose[2] and by Redington.[3] The extension of that analysis to include special trap distributions which avoid the conductivity limitation was reported by Rose and Lampert.[4] The emphasis in the present paper is on a physical understanding of these analyses.

A second, almost dramatic, illustration of the large departure from ideal performance encountered in insulating photoconductors, and of the difficulty of minimizing these departures, is given by the present status of television camera tubes of the Vidicon[5] type. The Vidicon makes use of a thin sheet of an insulating photoconductor contacted on one side by a transparent metal electrode and on the other side by a low-velocity scanning beam. A lifetime as large as 10^{-1} second can be tolerated and the free-carrier transit time through the thin sheet of photoconductor is in the order of 10^{-9} second. From these data, Equation (1) gives a photoconductive gain of 10.[8] Experimentally, many laboratories have attempted to make Vidicons with high gains. The highest reported values do not exceed unity. (The response time of a Vidicon for normal television use must not exceed 0.1 second.) The disparity of 10^8 between simple expectations as expressed by Equation (1) and experimental realization has been so great that it apparently has been largely overlooked.

A situation completely parallel to that of the Vidicon camera tube exists for the several photographic processes making use of an electrostatic charge pattern generated on an insulating sheet of photoconductor. These go by the names of Electrofax[6] and Xerography.[7] The

[2] A. Rose, "Maximum Performance of Photoconductors," *Helv. Phys. Acta*, Vol. 30, p. 242, 1957.

[3] R. W. Redington, "Maximum Performance of High-Resistivity Photoconductors," *Jour. Appl. Phys.*, Vol. 29, p. 189, February, 1958.

[4] A. Rose and M. A. Lampert, "Gain–Bandwidth Product for Photoconductors," *Bull. Amer. Phys. Soc.*, Vol. 3, p. 114, March 27, 1958; A. Rose and M. A. Lampert, "Photoconductor Performance Space-Charge Currents and the Steady-State Fermi Level," *Phys. Rev.*, Vol. 113, 1959, in press.

[5] P. K. Weimer, S. V. Forgue and R. R. Goodrich, "The Vidicon— Photoconductive Camera Tube," *RCA Review*, Vol. XII, p. 306, September, 1951.

[6] C. J. Young and H. G. Greig, "Electrofax—Direct Electro-Photographic Printing on Paper," *RCA Review*, Vol. 15, p. 469, December, 1954.

[7] R. M. Schaffert and C. D. Oughton, "Xerography: A New Principle of Photography," *Jour. Opt. Soc. Amer.*, Vol. 38, p. 991, December, 1948.

realized photoconductive gains are close to unity while the gains expected from Equation (1) are well over 10^6.

It is natural to expect that the explanation for departures of such magnitude and frequency as those just cited should be both simple and obvious. The simplicity is borne out in the analysis of the next section. However, the simplicity is deceptive because the analysis is neither obvious nor intuitive. The reason is that space-charge current flow forms the hinge of the argument and space-charge current flow in solids is a rarely discussed or encountered phenomenon.

ANALYSIS OF GAIN–BANDWIDTH PRODUCT

We consider here a homogeneous photoconductor having ohmic contacts and operated at fields below impact ionization. The ohmic contacts are logically essential for gains greater than unity.[*] Using Equation (2) as a starting point, we replace τ by the observable τ_0, the measured rise or decay time of photoconductivity, using the relation

$$\tau = \tau_0 \frac{n}{n_t} \tag{3}$$

where n is the density of free carriers and n_t is the density of free plus trapped charge in thermal contact with the free carriers. Thus, when n is altered by light, n_t must be altered by the same fraction. Equation (2) becomes

$$G \frac{1}{\tau_0} = \frac{n}{T n_t}. \tag{4}$$

In the absence of traps, $n/n_t \equiv 1$ and $\tau_0 \equiv \tau$. Now, if the following relations are introduced:

---

[*] Blocking contacts permit a gain of only unity at fields below impact ionization. Within this limitation, the gain–bandwidth product of a photoconductor using blocking contacts can be arbitrarily large, especially in the range of high insulation. (In the absence of light, a voltage applied across any material having blocking contacts tends to convert the material into an insulator.) A p-n junction operated in the back direction is a good example of the use of blocking contacts. The gain is well known to be close to unity, while the response time, given here by the transit time of the electrons and holes through the barrier, can be made almost arbitrarily small for small barrier thicknesses.

If, along with blocking contacts, the electric field is increased to the point of impact ionization, large gains as well as large gain–bandwidth products can be achieved. The control of impact ionization is, however, sufficiently delicate that no operating devices have yet made use of it.

$$\text{dielectric relaxation time} \equiv \tau_{\text{rel}} = \frac{10^{-12}K}{4\pi\sigma} = \frac{10^{-12}K}{4\pi\, ne\mu},$$

and $T = L^2/(V\mu)$, where L is the electrode spacing, then Equation (4) can be written as in Reference (4b),

$$G\frac{1}{\tau_0} = \frac{1}{\tau_{\text{rel}}}M; \qquad M \equiv \frac{N_A}{N_t}, \tag{5}$$

where N_A is the total number of electron charges on the anode and N_t is the total number of electron charges, free plus trapped, in thermal contact with the conduction band. In the discussion that follows, Equation (4) and/or Equation (5) are referred to frequently. Each has certain advantages in suggesting the physical phenomena involved.

We begin with a photoconductor operated at some low, finite voltage in the range where Ohm's law is valid. By Equation (4), the gain–bandwidth product increases linearly with increasing voltage because the transit time, T, decreases as V^{-1} and the free and trapped carrier concentrations remain invariant with applied voltage. Similarly, in Equation (5) only the factor N_A changes on the right hand side as the voltage is changed. Since N_A, the number of electron charges on the anode, is proportional to the applied voltage, so also is the gain–bandwidth product. It is clear also from Equations (4) and (5) that *at a given voltage in this low-voltage range*, the gain–bandwidth product increases as the density of traps is decreased and that this will continue only until the trap density is less than the free carrier density (see Figure 1).

Thus far, the discussion has reviewed phenomena that are well known and, so to speak, consistent with intuitive expectations. The critical question is now posed—*What is the maximum gain–bandwidth product that can be achieved as the voltage is increased?* If Ohm's law continued to be valid at high voltages, the gain–bandwidth product could obviously be increased indefinitely. Some new physical phenomenon must therefore intervene if the maximum gain–bandwidth product is to remain finite. The new phenomenon is the advent of space-charge current flow.* This occurs when the applied voltage is sufficient to inject more free carriers than are already present from

---

* It is assumed that space-charge current flow intervenes before dielectric breakdown. This can, in principle, be insured either by choosing a small enough electrode spacing or by choosing a material sufficiently free of traps.

either thermal or optical excitation.[8] By previous arguments, the total anode charge required to introduce space-charge currents must be equal to or greater than the total free plus trapped charge in thermal contact with the conduction band. If the anode charge is *just equal* to the free plus trapped charge at the onset of space-charge current flow, the value of M becomes unity (Figure 1) and, by Equation (5),

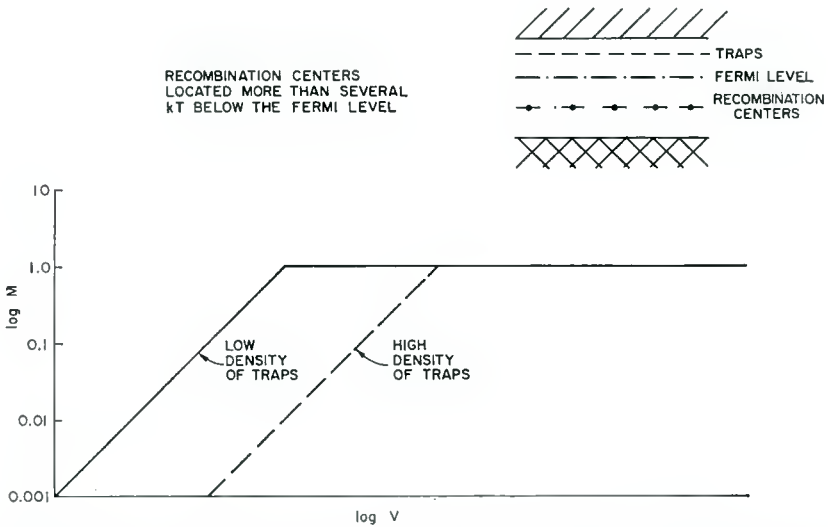$$G \times \frac{1}{\tau_0} = \frac{1}{\tau_{rel}}. \tag{6}$$

Fig. 1—M value versus voltage for recombination centers located more than several kT below the Fermi level. Note that while the trapping states are shown for convenience as a single level, they may also be distributed continuously in energy. The same is true for Figures 2-4.

If the anode charge *exceeds* that of the free plus trapped charge, $M > 1$ and significantly larger gain–bandwidth products can be achieved (Figure 2). We postpone for the moment the question of what happens at voltages beyond the advent of space-charge currents. The key problem then becomes: what are the relations between the anode charge and the trapped charge (in thermal contact with the conduction band) at the advent of space-charge current flow?

───────────

[8] R. W. Smith and A. Rose, "Space Charge Limited Currents in Single Crystals of CdS," *Phys. Rev.*, Vol. 97, p. 1531, March, 1955; A. Rose, "Space-Charge-Limited Currents in Solids," *Phys. Rev.*, Vol. 97, p. 1538, March, 1955; M. A. Lampert, "A Simplified Theory of Space-Charge-Limited Currents," *Phys. Rev.*, Vol. 103, p. 1648, September, 1956.

A completely general answer to the last question requires an under-
standing of the occupancy of the various discrete states in the forbidden
zone under conditions of illumination. The problem is extremely
involved, and there have been only a few attempts to treat it.[9] The
occupancy of a given state depends upon the rates of thermal generation
to each of the band edges, the rates of kinetic capture of electrons and
holes *from* each of the band edges, the capture cross sections of the
state for electrons and for holes, the location in energy of the state,
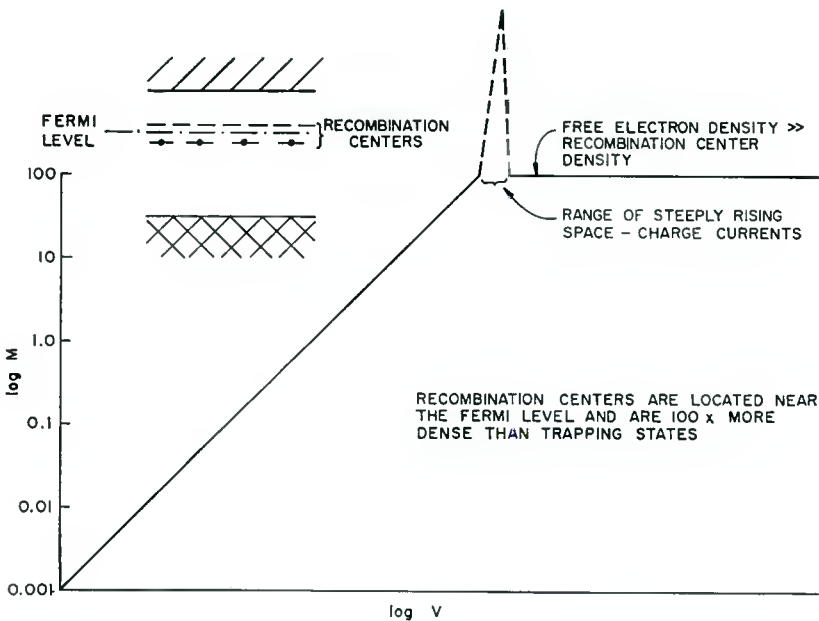the temperature, the rate of excitation of free pairs *and* the concen-

Fig. 2—*M* value versus voltage for recombination centers located
near the Fermi level.

[9] M. Schon, "Über die Strahlungslosen Übergange der Elecktronen im
Gitter der Kristal phosphore," *Tech. Wiss. Abh. der Osram*, Vol. 6, p. 49,
1953; I. Broser and R. Warminsky, "Statistisch-kinetische Theorie der
Lumineszenz und Electrischen Leitfahigkeit von Storstellenhalbleitern,"
Ann. der Phys., Vol. 16, p. 361, Heft 5/8, 1955; H. A. Klasens, "The Intensity
Dependence of Photoconduction and Luminescence of Photoconductors in
the Stationary State," *Jour. of Phys. and Chem. of Solids*, Vol. 7, p. 175,
November, 1958; R. H. Bube, "Analysis of Photoconductivity Applied to
Cadmium-Sulphide-Type Photoconductors," *Jour. of Phys. and Chem. of
Solids*, Vol. 1, p. 234, 1956; A. Rose, "Lifetimes of Free Electrons and
Holes in Solids," *Progress in Semiconductors*, John Wiley & Sons, New
York, N. Y., Vol. 2, 1957, p. 109; A. Rose, "Recombination Processes in
Insulators and Semiconductors," *Phys. Rev.*, Vol. 97, p. 322, January, 1955.

trations and properties of other discrete states in the forbidden zone. In the present paper, certain concepts will be utilized that have been developed in an earlier publication and that help to avoid some of the complexity. Moreover, emphasis will be placed on two limiting examples.

Consider an insulator in thermal equilibrium, as shown in Figure 3. Let a potential be applied across it sufficient to initiate space-charge
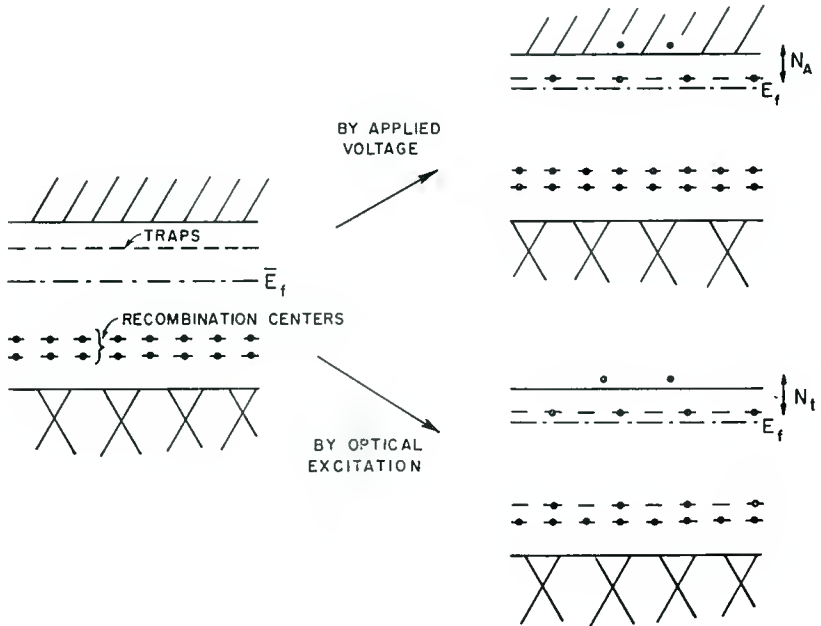

Fig. 3—Schematic representation of N_t and N_A showing that for shallow traps and deep lying recombination centers $N_t = N_A$. The figure shows the effect of raising the Fermi level by an applied voltage alone (space charge current flow) and by optical excitation alone.

current flow. This means that the Fermi level has been raised above its thermal equilibrium value by about kT. To do so, electrons must be injected into the insulator to double the free-electron density and to double the densities of electrons in all of the trapping levels down to the Fermi level. At the Fermi level, the number of electrons that must be supplied is approximately equal to the number of states lying within an energy slice kT wide. Usually this is more than half of the total number of electrons injected into the insulator.* For convenience

---

* This will be true for distributions in energy of discrete states that vary less rapidly than exp E/kT.

of argument only, these are considered to be the dominant states. At the voltage for the onset of space-charge flow, the number of electron charges on the anode (N_A of Equation (5)) is equal to the number of electron charges injected into the insulator and distributed throughout the volume.* This number, in turn, is given approximately by the total number of discrete states within kT of the Fermi level. The above equalities hold because the field lines from the anode end on charges distributed throughout the volume of the insulator rather than on charges near the cathode only.

It remains now to define N_t in Equation (5) in order to evaluate the critical factor $M \equiv N_A/N_t$. Suppose in the insulator just considered, the density of free electrons is doubled by optical excitation across the forbidden gap. The problem is to compute the total number of trapped electrons, N_t, that must be doubled along with the free carriers. While no general answer can be given, several significant possibilities can be outlined. The simplest is to assume that the free holes are rapidly captured by recombination centers located well below the Fermi level (Figure 1). Under these conditions, a doubling of the free-carrier density means that the Fermi level must be raised by kT in much the same way as described for the onset of space-charge currents. That is, all states within kT of the Fermi level must be filled with electrons. Thus, N_t and N_A are numerically equal. Physically, the rise time of the photocurrent is controlled by the same number of trapping states, N_t, as is the voltage at which space-charge flow is initiated.

The argument thus far has led to $M = 1$ at the onset of space-charge flow. The major assumption—capture of holes by recombination centers located more than several kT below the Fermi level—is one that can be expected to be frequently encountered and would therefore lead one to expect $M \leqq 1$ to be a common occurrence. At voltages below the onset of space-charge current, M is proportionately less than unity.

We now examine a second possibility, namely that the free holes are captured by recombination centers lying within kT of the Fermi level (Figure 4). It is immediately clear that doubling the free electron density (by optical excitation) certainly does not double the electron population of the recombination centers. On the contrary, the electron population in these states is *reduced* by optical excitation since the free electrons are effectively excited out of these centers. To compute N_t in this case, we seek states that are not recombination centers.

---

* At voltages below the onset of space-charge currents, the anode charge is also equal to the injected charge. At these lower voltages, however, the injected charge is localized near the cathode and has little effect on the conductance of the insulator.

Some of these may lie within kT of the Fermi level, while others will be found at various energy levels above the Fermi level. Those that are within kT of the Fermi level are counted in full; those lying ΔE above the Fermi level are first multiplied by the Boltzmann occupancy factor exp $(-\Delta E/kT)$ before being counted as part of N_t. In any event, the number N_t is less than the total number of states lying within kT of the Fermi level and therefore less than N_A. If the density of non-recombination centers (i.e., trapping states) can be made far less
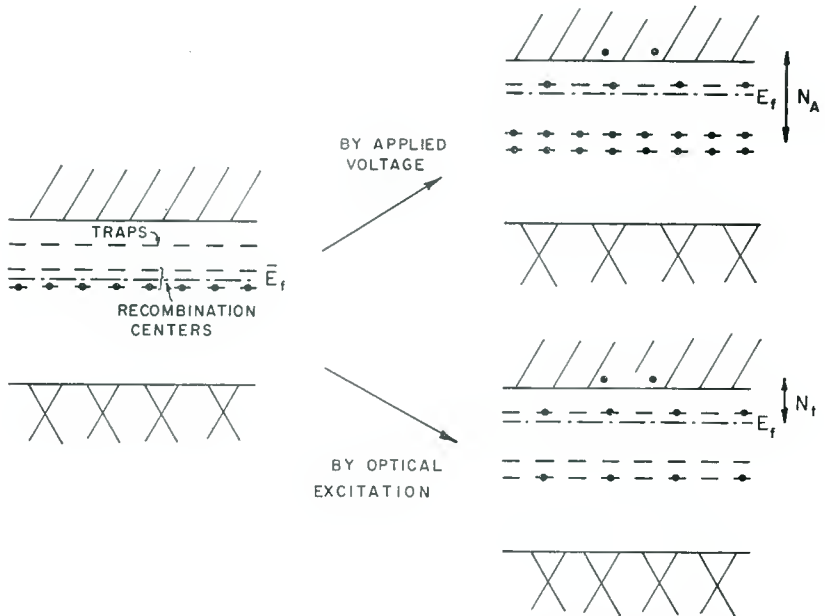

Fig. 4—Schematic representation of N_t and N_A showing that for recombination centers near the thermal equilibrium Fermi level $N_A \gg N_t$. The figure shows the effect of raising the Fermi by an applied voltage alone (space charge current flow) and by optical excitation alone.

than the density of recombination centers, the value of M can be made correspondingly large.*

While the above arguments have been carried out for photocurrents in the neighborhood of the thermal equilibrium dark current, they are approximately valid also for small changes in photocurrents in the neighborhood of a large d-c bias photocurrent. In this case, N_t is

---

* When $M > 1$, it is possible that the time required for readjustment at the ohmic contact may limit M to a value less than that to be expected from the ratio N_A/N_t discussed in this paper. See M. A. Lampert and A. Rose, "Transient Response of an Ohmic Contact," *Phys. Rev.*, Vol. 113, 1959, in press.

computed from trapping states near and above the steady-state Fermi level set by the d-c bias light. $N_{,1}$, however, includes in addition all the states between the dark Fermi level and the steady-state Fermi level. The result is that M values modestly greater than unity can be achieved even for deep-lying recombination centers. For a relatively uniform distribution of traps in energy, and for voltages close to the threshold of space-charge currents, the M value is $\Delta E/kT$ where ΔE is the separation of the two Fermi levels. The new steady-state Fermi level could also have been produced by an applied voltage in the dark, giving rise to space-charge currents, rather than by a d-c bias light. However, insofar as space-charge currents are usually a steeply rising function of the applied voltage, this is not a convenient mode of operation.

Device Implications

Equation (6) gives the maximum performance to be expected in many, if not most, photoconductors. For some specialized distributions of states in the forbidden zone this performance can be exceeded by the factor M as already discussed. While the immediate discussion is confined to the limitations given by Equation (6), the enhancement factor, M, can readily be borne in mind.

Consider a simple CdS photoconductor for which a gain of 10^5 and a response time of 10^{-3} second are commonly reported values. Equation (6) indicates that to achieve these values, the dielectric relaxation time of the CdS must be shorter than 10^{-8} second or, what is equivalent, the conductivity of the CdS must be greater than 10^{-4} (ohm-cm)$^{-1}$. Suppose that this conductivity was brought about by a bias light and that the dark conductivity was negligible. If the same CdS cell is operated at very low lights in the neighborhood of 10^{-10} (ohm-cm)$^{-1}$ the corresponding dielectric relaxation time is 10^{-2} second. Equation (6) now indicates that if the gain is still 10^5, the response time of the cell will be equal to or greater than 10^3 seconds or 10^6 times larger than its high-light value.

The data reported by H. B. DeVore[10] and R. W. Smith[11] bear out the marked drop in performance of CdS and CdSe photocells as the conductivity under the operating conditions is lowered. Even so, in both cases, M values significantly greater than unity were observed at

[10] H. B. DeVore, "Gains, Response Times and Trap Distribution in Powder Photoconductors," *RCA Review*, Vol. XX, p. 79, March, 1959.

[11] R. W. Smith, "Properties of Deep Traps Derived from Space-Charge Current Flow and Photoconductive Decay," *RCA Review*, Vol. XX, p. 69, March, 1959.

very low illumination. These data are consistent with the assumption of recombination centers (in excess of trapping states) located at the thermal equilibrium position of the Fermi level. For higher positions of the Fermi level corresponding to higher steady-state illuminations, the trap densities near the Fermi level approach or exceed the density of recombination centers and M approaches unity.

The application of Equation (6) to the Vidicon leads immediately to a maximum gain of unity—a value not yet exceeded in practice. From Equation (6),

$$ G = \frac{\tau_0}{\tau_{rel}} . $$

The smallest value of τ_{rel} is set at 10^{-1} second by the condition that charges be stored for a picture time. The largest value of τ_0 is also set at 10^{-1} second by the condition that the photoconductivity excited in one picture time not overlap into succeeding pictures. These two limiting time constants give a gain of unity consistent with experience.

The highest reported gains for the photographic processes Electrofax and Xerography are close to unity. Since both processes currently make use of blocking contacts, a gain in excess of unity is not to be expected. On the other hand, if gains greater than unity could be achieved by the use of ohmic contacts, they undoubtedly would have been reported. But, by Equation (6), a gain of only unity is to be expected insofar as $\tau_0 \leqq \tau_{rel}$. If $\tau_0 > \tau_{rel}$, the operating gain remains unity because the charged layer is discharged by light before full use is made of the slowly rising photoconductivity.

In the case of light amplifiers of the type described by Nicoll and Kazan,[12] photoconductive gains considerably greater than unity have been achieved. The permissible response time for reproducing motion is again a picture time or about 0.1 second. The dielectric relaxation time, however, should be approximately the reciprocal of the frequency of a-c voltage used to operate the amplifier. For a-c frequencies of 10,000 cycles per second, a photoconductive gain of 10^3 is to be expected from Equation (6).

In all of the devices cited, the value of M was assumed to be unity. Such an assumption accounts for most of the reported performances. Higher performance can, nevertheless, be achieved providing the distribution and character of the discrete states in the forbidden zone can be properly tailored as discussed in the preceding section.

---

[12] B. Kazan and F. H. Nicoll, "Electroluminescent Light Amplifying Picture Panel," *Proc. I.R.E.*, Vol. 43, p. 1888, December, 1955.

PROPERTIES OF DEEP TRAPS DERIVED FROM SPACE-CHARGE-CURRENT FLOW AND PHOTOCONDUCTIVE DECAY*

By

R. W. SMITH

RCA Laboratories,
Princeton, N. J.

Summary—The states deep within the forbidden band of an insulator can be scanned either by moving the Fermi level by voltage injected carriers or by moving the Fermi level by optically excited carriers. The current–voltage characteristics of a single crystal of CdS have been measured both in the dark and at different light levels over an extended range of voltages. From these data along with photoconductive gain and decay time measurements the performance (gain–bandwidth) of the photoconductor can be evaluated and some of the characteristics of the deep states determined. For the crystal measured, the dark current varied as V^2 in the low-voltage range and then increased steeply with V. This corresponds to the Fermi level starting in a trap-free region at ~0.9 ev below the conduction band and then sweeping through a single trapping level N_t of density ~10^{14} per cubic centimeter at 0.8 ev. At low light levels the photoconductor M value attained a maximum value of 50 at the onset of space-charge current flow. As the Fermi level was raised by raising the light level the maximum M obtained decreased to approximately 5. M values greater than unity indicated that a predominant number of states at the 0.8 ev level were not functioning as traps but as recombination centers. The decay time, at a fixed light level, increased by as much as a factor of 100 as the voltage was increased to the onset of space-charge currents. This indicates that the photoconductive performance is determined by the states in the region of the contact rather than by the states in the body of the insulator.

INTRODUCTION

THE Fermi level can be moved through a range of states within the forbidden band of an insulator by injecting excess (space-charge-limited) electrons into the crystal with an applied voltage or by optically exciting carriers in the crystal. Characteristics of the states scanned by the Fermi level can be obtained from an analysis of the current–voltage characteristics and from the photoconductive response-time data for perturbations in the current density. The purpose of this paper is to describe measurements on a single crystal of CdS in which deep-lying states are examined by this technique. A single crystal is used with a geometry designed to minimize surface

---

effects and to permit accurate determination of the amount of light absorbed.

With the crystal in the dark, and for the range of current covered, the current is carried exclusively by electrons injected into the crystal from the cathode. When only a single type of carrier is present (in contrast to the case with optical excitation) there is a minimum departure from thermodynamic equilibrium. This means that in the steady-state the free carriers are in thermal equilibrium with *all* the trapping states and that the distribution of electrons in traps and in the free states can be described by a steady-state Fermi level. It has already been demonstrated[1] that the density and distribution of trapping states can be obtained from simple current–voltage data. New data obtained by this method indicating the existence of a *single* level deep in the forbidden zone are described.[2]

Under optical excitation, both electrons and holes are generated and recombine through states within the forbidden band. The distribution of electrons into these states now becomes more complex and is influenced by the kinetics of recombination as well as by thermal exchange. The performance, i.e., gain–bandwidth product of the crystal when used as a photoconductor, is sensitive to this distribution. Moreover, it has only recently been realized that space-charge injection can be a factor limiting the operation of a photoconductor. One of the purposes of this paper is to examine the photoconductor in terms of an expression for performance, which includes space-charge injection, developed by Rose and Lampert.[3] It is shown that correlated data obtained from a single crystal both in the dark and with irradiation provide a means for identifying a set of levels as either trapping states or recombination states.

EXPERIMENTAL PROCEDURE

The measurements were made on a photosensitive CdS:Cu:Cl crystal. An opaque indium electrode was located on one side of the

---

[1] R. W. Smith and A. Rose, "Space-Charge-Limited Currents in Single Crystal of Cadmium Sulfide," *Phys. Rev.*, Vol. 97, p. 1531, March, 1955; A. Rose, "Space-Charge-Limited Current in Solids," *Phys. Rev.*, Vol. 97, p. 1538, March, 1955; M. A. Lampert, "Simplified Theory of Space-Charge-Limited Currents in an Insulator With Traps," *Phys. Rev.*, Vol. 103, p. 1648, September, 1956; M. A. Lampert, A. Rose, and R. W. Smith, "Space-Charge-Limited Currents as a Technique for the Study of Imperfections in Pure Crystals," *Jour. of Phys. and Chem. of Solids*, to be published.

[2] R. W. Smith, "Determination of a Discrete Trapping Level in CdS by Space-Charge Current Measurements," *Bull. Amer. Phys. Soc.*, Vol. 3, p. 129, March 27, 1958.

[3] A. Rose and M. A. Lampert, "The Gain–Bandwidth Product for Photoconductors," *RCA Review*, Vol. XX, p. 57, March, 1959; also "Photoconductor Performance, Space-Charge Currents and the Steady-State Fermi Level," *Phys. Rev.*, Vol. 113, 1959 (in press).

crystal and a semitransparent indium electrode of known transmission was located opposite it on the other side of the crystal. The sample was excited with monochromatic light corresponding to the peak of photoresponse. With this arrangement, surface effects are minimized and conductivity within the body of the crystal is measured. The electrodes were evaporated on to the crystal and there was no additional processing. The range over which ohmic behavior is obtained confirms that this simple procedure is sufficient to give ohmic electrodes.

The current was measured with a vibrating-reed electrometer. The most troublesome feature of the measurements is the long time required for the current to come to equilibrium each time it is perturbed, particularly at low light levels. To obtain reproducible results, the crystal was placed in the dark for several days with no voltage applied. Voltage was then applied in small increments with sufficient time being allowed for the current to come to equilibrium. In the low light region, approximately a day was required for each point. A useful identifying property of single-injection space-charge current is an overshoot in current followed by decay to a steady state. This occurs each time the voltage is raised. At very high lights and for voltages giving currents well up on the steep curve of Figure 1, the current continues to rise at a fixed voltage. This probably indicates the onset of double injection with holes tunneling in from the anode.

Dark Characteristic

The curve I_0 of Figure 1 is the thermal equilibrium dark current as a function of voltage. There are several distinct branches to the curve; an Ohms-law section at low voltages, a square-law branch, and a section where the current varies very steeply with voltage. The Ohms-law section may be associated with the residual (donor) conductivity of the crystal. At this time, the data is not sufficiently complete to distinguish this from residual photoconductivity or from spurious leakage. The low-current measurement, however, does indicate that the dark resistivity is $\geqslant 10^{15}$ ohm-centimeter, and that the Fermi level is initially located ~ 0.95 ev below the conduction band.

The variation of the current with V^2 and then as a steep function of V has a particularly simple and direct interpretation in terms of the distribution of states through which the Fermi level moves. The number of electrons injected into the crystal is proportional to the applied voltage. While most of these electrons fall into trapping states, a small fraction remain free and account for the current measured. The density of electrons in the conduction band determines the location

of the steady-state Fermi level among the traps. As the voltage is increased the precise value of the current depends on the specific density and distribution of states through which the Fermi level moves. A single discrete level of density, N_T, is shown at E_T in Figure 2. As long as the Fermi level is below E_T and moves in a region in which
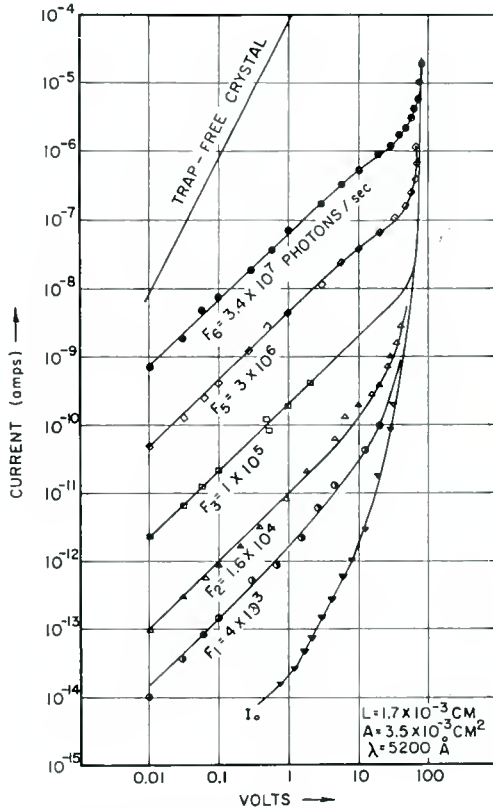

Fig. 1—Current voltage characteristics. I_0 is the dark characteristic. The photo-current for different light levels is shown by the F curves. The trap-free crystal curve is the current expected from the crystal if there were no traps present.

the density of background states is negligible in comparison with N_T, the ratio, θ, of free electrons, n, to the number in the N_T states, n_t, is constant and independent of V.[1] The current density in this region is

$$J = 10^{-13} \, \mu \kappa \theta \, \frac{V^2}{L^3}, \; (\text{amp/cm}^2) \tag{1}$$

where all the quantities are in practical units and

$$\theta = \frac{n}{n_t} = \frac{N_c}{2N_T} e^{-\frac{E_T}{kT}}, \tag{2}$$

μ is the electron mobility, κ is the dielectric constant, and $N_c = 2 (2\pi m^* kT/h^2)^{3/2}$ is the density of states at the bottom of the conduction band. The density of states effective mass is m^*. To determine θ, the current measured in the V^2 region is compared with the current expected from the trap-free crystal (the upper left line in Figure 1)

Fig. 2—Energy band structure of CdS with a single discrete level deep within the forbidden band.

From the measured θ and N_T (see below), the discrete level is calculated to be 0.8 ev below the conduction band.

The extremely steep part of the curve is characteristic of injected electrons filling most of the discrete level. It has been shown that the trap density, N_T, is given directly by the voltage, V_{TF}, at the steep part of the curve.

$$eN_T = \frac{10^{-12}\kappa}{2\pi L^2} V_{TF}, \tag{3}$$

where L is the thickness of the crystal. From V_{TF} in Figure 1, $N_T = 3 \times 10^{14}$ per cubic centimeter is computed. The Fermi level is centered at the trap level at the voltage $2V_{TF}/3$. The location of the

trap, E_T, is calculated from the carrier density corresponding to $2V_{TF}/3$. From Figure 1 this also gives $E_T = 0.8$ ev.

PHOTOCURRENT

The V–I characteristics for photocurrent at 5 different light levels are shown in Figure 1. The current in each case was extended well up into the space-charge region. The time, τ_0, for the current to decay to one-half its initial value when the light is removed, was measured at each experimental point shown on the curve. From this data the gain, $G = (\Delta I/eF)_V$, the bandwidth, $1/\tau_0$, and the relaxation time

$$\tau_{RC} = (RC)_V = \frac{V}{I} \frac{KA}{L} \tag{4}$$

were recorded as a function of voltage. Here K is a constant which depends on the geometry of the sample and A is the electrode area. The performance of the photoconductor can now be evaluated as the Fermi level sweeps through the same range of energy covered by the dark current data. The expression developed by Rose and Lampert[3] for the gain–bandwidth product for a photoconductor is

$$G \frac{1}{\tau_0} = \frac{1}{\tau_{RC}} M. \tag{5}$$

M was determined as a function of voltage for the five different light levels. Figure 3 shows τ_0 as a function of V. Figure 4 shows M versus V for two light levels. In Table I the M values for all the light levels are included.

Significant features of the photoconductivity data include:

1. The wide range over which the photocurrent follows Ohms law and the smooth transition into the steep space-charge-limited region. Note that at the higher light levels, Ohm's law is observed for the same range of voltages for which the dark current is non-ohmic and space-charge limited.

2. The M value increases with voltage to as high as 50 near the space-charge region, at low lights. As the Fermi level moves up through the forbidden zone as the light level is raised, the maximum M decreases to a value nearer unity.

3. The decay time, τ_0, for a fixed light level is essentially constant at low voltages and then increases, by as much as a factor of
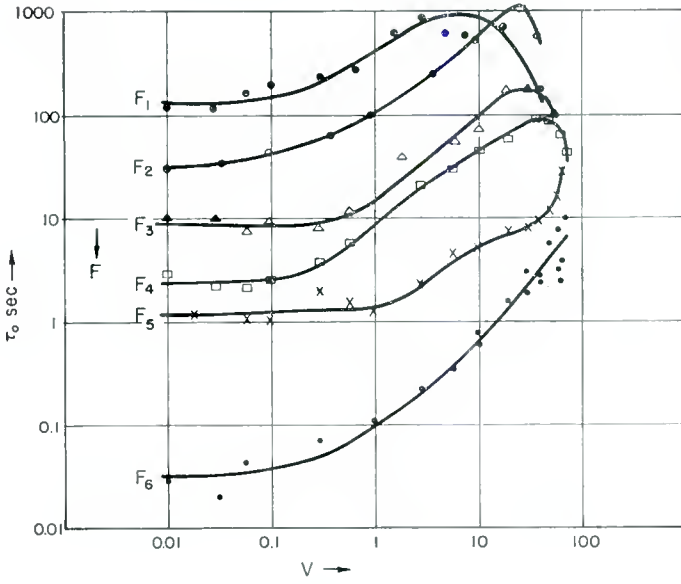
Fig. 3—Photoconductive decay time, τ_0, as a function of voltage for different light levels.
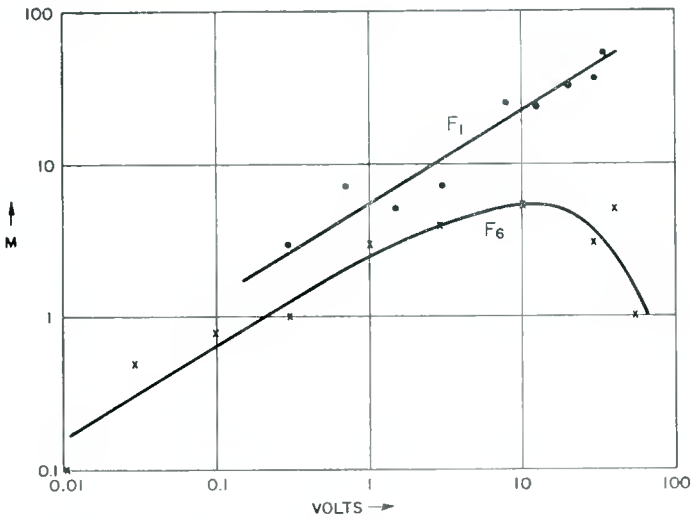

Fig. 4—Variation of M value with voltage for two extreme light levels.

Table I

V	F_1	F_2	F_3	F_5	F_6
			M		
0.01	..	0.2	..	0.02	0.1
0.03	..	0.7	..	0.05	0.5
0.1	..	2	1	0.35	0.8
0.3	3	..	..	0.58	1
1.0	5	7	5	3	3
3	7	..	..	5	4
10	25	13	13	7	5
20	32	..	..	..	..
25	..	18	..	..	..
30	36	..	16	15	3
35	51	44	..	..	..
40	..	..	18	16	5
50	..	..	26	..	1
60	..	..	39	..	3
65	..	..	44	..	2

100 with voltage, to a maximum near the onset of the steep rise in current.

DISCUSSION

Of particular interest is the value of M and its significance as regards the nature of deep-lying states at 0.8 ev below the conduction band. These states and their density, 3×10^{14} per cubic centimeter, were determined from the dark-current data. The space-charge-limited current curve describes the density of all states and does not differentiate between trapping states and recombination states. Optical excitation is needed to make this distinction. Under optical excitation those states whose occupancy is determined by thermal exchange with the carriers in the conduction band are called trapping states, while those states whose occupancy is determined by kinetic processes of recombination are called recombination states.

From an earlier analysis[3] it was shown that the M value could exceed unity only if the density of recombination states near the Fermi level exceeded the density of trapping states. The large M values recorded above indicate that the states at 0.8 ev are, indeed, recombination states.

Again from an earlier analysis,[4] it was pointed out that once the M value exceeds unity, there are two possibilities for determining its actual value. The M value should take on the smaller of two values; the ratio of number of electron charges on the anode to the number of

---

[4] M. A. Lampert and A. Rose, "Transient Response of an Ohmic Contact," *Phys. Rev.*, Vol. 113, 1959 (in press).

trapping states in the body of the crystal or the ratio of the number of electron charges on the anode to the number of electron charges in a Debye length* near the cathode (see Figure 5). In the former, the M value and response time are volume controlled, and in the latter, contact controlled. To differentiate between these two possibilities, use is made of the fact that the volume-controlled response time should not be voltage dependent, whereas the contact-controlled response time
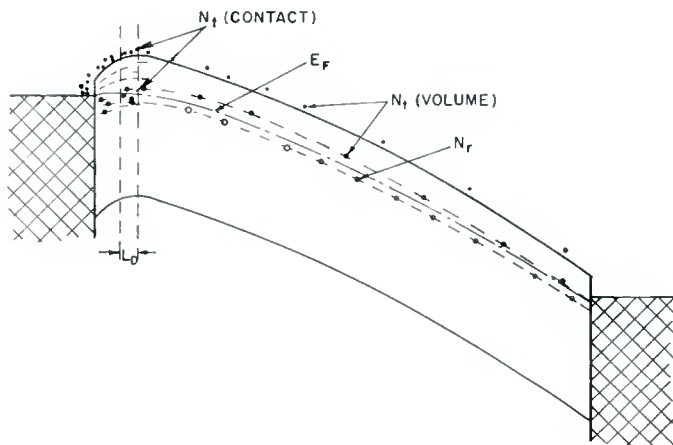

Fig. 5—Diagram indicating the distribution of centers and space-charge for body and for contact controlled operation. The contact-controlled photo-response time is determined by the number of excess charges within a Debye length near the contact and at the energy maximum. The recombination states within the Debye length are included here because they contain excess charge whose lines of force must shift from being directed towards the cathode to being directed towards the anode when the photocurrent is increased by a factor of two. (See ref. 3)

should increase with voltage if the trap density increases as one approaches the surface—a reasonable expectation. The data show a marked increase of response time with increasing voltage, namely some 10–100 fold. It must be concluded then that the response time and M values are contact controlled.

Finally, some quantitative estimate of the maximum M value to be expected can be made. If the recombination state density were assumed to be constant out to the surface and if this were the dominant state density, the maximum M value (achieved at the onset of space charge

---

* The Debye length is evaluated at the energy maximum and is given by $L_{l\!\!/} = 7(\kappa T/N_T)^{1/2}$ centimeters.

limited current flow) would be given by the ratio of the electrode separation to a Debye length.[4] For the recombination state density of 3×10^{14} per cubic centimeter, the Debye length is 2×10^{-5} centimeter. This value combined with the electrode separation of 10^{-3} centimeter gives a maximum M value of 50. This computed value compares favorably with the experimental value of 50 achieved at low lights where the recombination state density was most likely to be dominant near the surface. At higher light intensities the maximum M value observed experimentally was ~5. Here we can expect the number of electron charges in a Debye length to be larger because the potential minimum (energy maximum) for electrons lies closer to the surface and closer to the Fermi level. The trap density is likely to be greater near a physical surface and is also likely to be greater near the conduction band (shallow traps). The Debye length is measured at the potential minimum and the number of electron charges in a Debye length increases as the square root of the trap density.

The increasing density of traps near the surface is also reflected in the behavior of the response time. At low lights, where surface and volume densities are comparable, the response time increases only by a factor of 3 as the voltage increased from M ~1 to M(max). At high lights, the response time in the same range increases by a factor of about 30. The contact-controlled response time is also proportional to the number of electron charges in a Debye length. As long as the response times and the M values are dominated by traps at or near the contact, information about the density of volume traps lying above the recombination centers is masked.

Acknowledgment

GAINS, RESPONSE TIMES, AND TRAP DISTRIBUTIONS IN POWDER PHOTOCONDUCTORS*

By

HENRY B. DeVORE

RCA Laboratories,
Princeton, N. J.

Summary—Gain and photocurrent decay measurements have been made for CdS and CdSe powder photoconductors. These data are compared with the analytical expression derived by Rose and Lampert, relating gain–bandwidth product to dielectric relaxation time. The gain–bandwidth product is, in the case of low illumination levels, considerably larger than the reciprocal relaxation time. This can formally be accounted for on the assumption of an admixture of recombination states and trapping states near the Fermi level.

Trap distributions have been derived from photocurrent decay curves. These generally show a nearly exponential decrease of trap density with increasing trap depth below the conduction band.

INTRODUCTION

THIS paper presents the results of measurements of photoconductive gains, response times, and trap distributions for cadmium sulfide and cadmium selenide powder photoconductors. A. Rose and M. Lampert[1] have analyzed the gain and response-time relationships for photoconductors and have shown that

$$G \frac{1}{\tau_0} = \frac{1}{\tau_{rel}} M, \tag{1}$$

where G is the gain, τ_0 is the time constant for build-up or decay of the photoconductive current, τ_{rel} is the dielectric relaxation time (RC) under the conditions of illumination, and M is a factor which should increase with applied voltage to a maximum value which should, in general, be approximately unity, but may exceed unity with some special forms of trap distribution.

It has been found, particularly in the case of cadmium selenide, that values of M exceeding unity by some two orders of magnitude may be observed. A study of trap distributions was undertaken in search

---

* Manuscript received November 4, 1958.
[1] A Rose and M. Lampert, "The Gain-Bandwidth Product for Photoconductors," *RCA Review*, Vol. XX, p. 57, March, 1959.

of an explanation for the large M factors. While the results failed to provide the explanation sought, the method of trap distribution measurement, based on photocurrent decay curves, is of sufficient interest to warrant a description.

MATERIALS

The photoconductive materials studied were CdS and CdSe powders, doped with Cu and Cl to achieve high photoconductive sensitivity and low dark current.[2] These were used either as powders embedded in a plastic binder or as a thin layer sintered to the surface of a ceramic disc.[3] The electrode structure is a pair of coplanar conducting sheets arranged to provide surface electrodes separated by a gap 0.5 millimeter wide and 5 millimeters long. For the sintered photoconductor layers, the electrodes were silver paint coated on the photoconductor surface. For the powders, the electrodes were conductive strips on a glass base. The photoconductive powder was dispersed in a 1 per cent solution of ethyl cellulose in diacetone alcohol, the concentration being such as to form a freely flowing mixture. A drop of this was placed over the gap and dried. In both cases, illumination was through the gap.

The powder materials are generally characterized by nonlinear behavior.[4] The photocurrent for a powder in an organic binder varies approximately as the fourth power of the applied voltage. On the other hand, for sintered layers, the photocurrent-voltage relationship is substantially linear. The fourth-power behavior for powders results from the effect of the numerous particle contacts. Similar behavior has been reported for other powdered semiconductors.

The photocurrent–light-intensity relationship varies from somewhat less than linear at the higher intensities to considerably greater than linear slope at low intensities. Figures 1 and 2 show representative photocurrent–light-intensity cureves.

The photoconductive powders have a spectral response which is considerably broader than the corresponding curve for a single crystal and which is extended towards longer wavelengths. The sintered layers have narrower spectral response curves than the corresponding powders,

---

[2] S. M. Thomsen and R. H. Bube, "High Sensitivity Photoconductor Layers," *Rev. Sci. Inst.*, Vol. 26, p. 664, July, 1955; The powders discussed herein were experimental types. The powders designated CdS 1424A and 1434A are similar to materials marketed by the RCA Electron Tube Division as Developmental Types F2103 and F2111.

[3] The sintered samples were supplied by R. W. Engstrom of the RCA Electron Tube Division, Lancaster, Pa.

[4] F. H. Nicoll and B. Kazan, "Large-Area High-Current Photoconductive Cells Using CdS Powder," *Jour. Opt. Soc. Amer.*, Vol. 45, p. 647, August, 1955.

but still much wider than those for single crystals. Figure 3 shows representative spectral response curves for these photoconductors.[5]

GAIN, RESPONSE TIME, AND RELAXATION TIME

The gain of a photoconductor is defined as the number of charge carriers crossing the gap between electrodes per photon absorbed in the photoconductor in the gap.
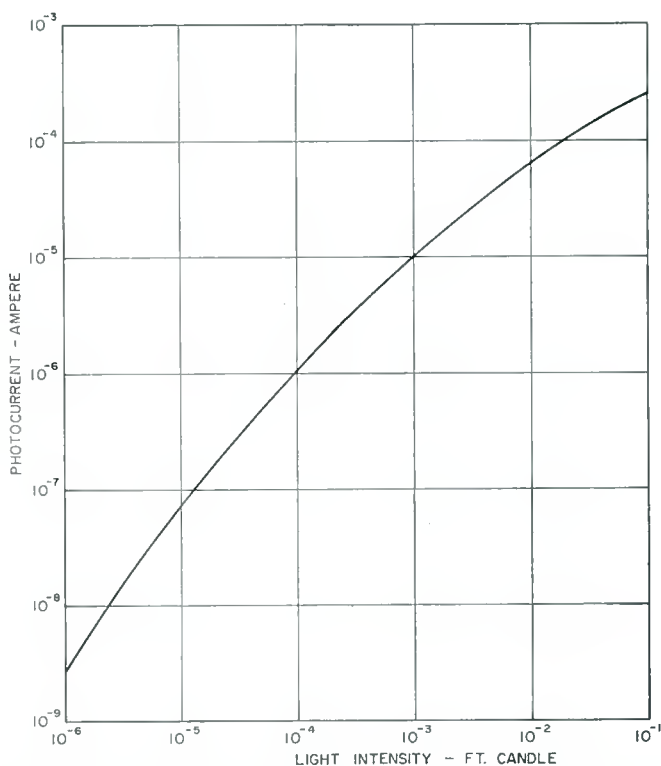

Fig. 1—Photocurrent versus light intensity — CdS powder 1424A.

For an illumination of L foot candles at the sample, the light flux falling on the electrode gap is $2.7 \times 10^{-5} L$ lumens per second and the effective absorbed photon flux is $\phi_{\text{eff}} = 2.7 \times 10^{-5} LP$, where P is a photon efficiency factor which can be obtained from the following expression:

$$P = \int W_\lambda \, S_\lambda \, \Psi_\lambda \, d\lambda,$$

---

[5] The spectral response measurements were made by W. F. Davidson.

Fig. 2—Photocurrent, versus light intensity — CdSe sintered.

Fig. 3—Spectral response curves for powder photoconductors.

where W_λ is the radiant power from the light source at wavelength λ,

S_λ is the photoconductive response of the sample at wavelength λ,

Ψ_λ is the photon flux equivalent at wavelength λ.

If the photocurrent corresponding to this illumination is i_L amperes, corresponding to $n_c = 6.2 \times 10^{18} \, i_L$ electrons per second crossing the electrode gap, then the gain is

$$G = \frac{n_e}{\phi_{\text{eff}}} = 2.3 \times 10^{23} \frac{i_L}{PL}. \tag{2}$$

The response time τ_0, corresponding to light intensity L, is obtained by measuring the time required, after illumination is cut off, for the photocurrent to decay from the equilibrium current i_L at this light intensity to $1/e$ of this value.

The relaxation time $\tau_{\text{rel}} = R_L C$, where R_L is the resistance of the illuminated photoconductor and C is the capacitance of the photoconductor in the gap to its surroundings. C was determined by calculation and by experiment to have the value 0.5 micromicrofarad for the gaps considered, and $R_L = V/i_L$, with V being the voltage applied.

Measurements were generally made with 400 volts applied across the gap. Higher potentials entailed risk of permanent damage by dielectric breakdown. At the operating potential, the dark current did not show the steep rise characteristic of space-charge-limited currents.[6] The Rose–Lampert[1] analysis indicates that for simple trap distributions, the M values should be less than unity for applied voltages below that required for space-charge-limited currents. They have proposed an explanation of M values in excess of unity on the assumption of a concentration of recombination states in the neighborhood of the Fermi level.

Combining Equations (1) and (2) with the expression for τ_{rel},

$$M = \frac{4.6 \times 10^{13}}{P L \tau_0}. \tag{3}$$

For the samples studied, representative values of P and the corresponding formulas for G and M are given in Table I. Since the

---

[6] R. W. Smith and A. Rose, "Space-Charge-Limited Currents in Single Crystals of Cadmium Sulfide," *Phys. Rev.*, Vol. 97, p. 1531, March 15, 1955.

sintered layers of both CdS and CdSe have narrower spectral responses than the corresponding powders, the photon efficiencies are higher.

Table II shows some measurements of photocurrent and decay time constants for samples of CdS and CdSe, together with the corresponding calculated values of G and M. It may be noted that different samples of CdS powder have variations as high as 15:1 in photocurrent and gain, but the time constants and the M factors only vary over a 2:1 range.

Generally as the illumination intensity is decreased, the time constant increases, but at a slower rate, so that the M factor increases. This is shown in Table III. For sintered CdSe, it was found that the M factor was in excess of unity for light intensities below 10^{-2} foot candle, and considerably in excess of unity at very low light levels. Rose and Lampert have discussed models of photoconductors which may be expected to show large M-factor values for some ranges of illumination.

Trap Distributions

The study of the decay of photocurrents provides a means of determining the distribution in energy of trap densities in a photoconductor. This method was outlined by Rose[7] and used by Gerritsen, Ruppel, and Rose[8] in a study of zinc oxide. A generically related method for trap distribution studies, based upon measurement of response to fluctuating light, was described by Niekisch.[9]

The interpretation of the photocurrent decay measurements is as follows: With the photoconductor in equilibrium with the light, a certain density of free electrons will exist in the conduction band. This density may be calculated from the conductivity. These electrons will be in thermal equilibrium with a much larger number of electrons in traps which will be filled up to a limit depending on the conduction band density. This limit will define the position of a quasi-Fermi level. There is, then, an equilibrium among electrons changing between conduction band and traps, electrons raised to the conduction band by light, and electrons falling into the large number of deep-lying recombination centers. When light is cut off, the electrons continue to

[7] A. Rose, "An Outline of Some Photoconductive Processes," *RCA Review*, Vol. XII, p. 362, September, 1951.

[8] H. J. Gerritsen, W. Ruppel, and A. Rose, "Photoproperties of Zinc Oxide with Ohmic and Blocking Contracts," *Helv. Phys. Acta*, Vol. 30, p. 504, November, 1957.

[9] E. A. Niekisch, "Wechsellichtmessungen an photoleitendem CdS als Methode zur Bestimmung von Beweglichkeit und Haftstellenverteilung," *Ann. der Phys.*, Vol. 15, p. 279 and 288, 1955.

Table I—Photon Efficiencies, P, Gains, G, and M-factors for Powder Photoconductors

	P (PHOTONS/SEC./FT.CANDLE)	G	M
CdS SINTERED	3.7×10^{15}	$6.2 \times 10^7 \; i_L/L$	$\dfrac{1.2 \times 10^{-2}}{\tau_0 L}$
CdS POWDER	1.2×10^{16}	$2.0 \times 10^7 \; i_L/L$	$\dfrac{4.0 \times 10^{-3}}{\tau_0 L}$
CdSe SINTERED	1.5×10^{16}	$1.6 \times 10^7 \; i_L/L$	$\dfrac{3.2 \times 10^{-3}}{\tau_0 L}$
CdSe POWDER	2.8×10^{16}	$8.4 \times 10^6 \; i_L/L$	$\dfrac{1.7 \times 10^{-3}}{\tau_0 L}$

Fig. 4—Initial photocurrent decay — CdS powder 1424A.

Table II—Gains and M-factors for Powder Photoconductors
$L = 0.1$ foot-candle, $V = 400$ volts.

	CdS POWDERS				CdS SINTERED	CdSe POWDER	CdSe SINTERED
	1007	27-2	1424A	1434A			
i_L (μA)	69	29	430	280	75	22	750
τ_0 (SEC.)	0.86	0.65	0.39	0.80	0.090	0.034	0.056
G	1.4×10^4	5.8×10^3	8.6×10^4	5.6×10^4	4.7×10^4	1.8×10^3	1.2×10^5
M	0.047	0.062	0.10	0.050	14	0.49	0.57

Table III—Gains and M-factors for Powder Photoconductors — Variation with Illumination Level

	FT. CANDLE	10^{-1}	10^{-2}	10^{-3}	5×10^{-5}	2×10^{-6}
CdS POWDER 1007	i_L (μA)	60	6.0	0.43		
	τ_0 (SEC.)	0.85	2.6	10.8		
	G	1.2×10^4	1.2×10^4	8.6×10^3		
	M	0.060	0.19	0.46		
CdS SINTERED II	i_L (μA)	75	3.6	0.21		
	τ_0 (SEC.)	0.090	0.40	5.0		
	G	4.7×10^4	2.2×10^4	1.3×10^3		
	M	1.4	3.1	2.5		
CdSe POWDER IID	i_L (μA)	22	0.75			
	τ_0 (SEC.)	0.034	0.082			
	G	1.8×10^3	6.3×10^2			
	M	0.49	2.0			
CdSe SINTERED-I	i_L (μA)	750	95	5.5	0.20	0.009
	τ_0 (SEC.)	0.056	0.11	0.39	4.3	29
	G	1.2×10^5	1.5×10^5	8.8×10^4	5.8×10^4	6.6×10^4
	M	0.57	2.9	8.2	15	55

Fig. 5—Photocurrent decay — CdS powder 1424A.

Table IV—Formulas for Calculation of Trap Distribution

EXCITATION DENSITY = f_L = $1.08 \times 10^{-3} \dfrac{LP}{d}$	FOR STANDARD GAP $\begin{cases} \ell = 0.05 \text{ cm} \\ w = 0.5 \text{ cm} \end{cases}$ AND $V = 400$ VOLTS, $\mu = 10$ cm²/VOLT SEC.
CONDUCTION BAND DENSITY = n_o = $\dfrac{i_L}{e\mu V} \cdot \dfrac{\ell}{w_d}$	$= 1.6 \times 10^{14} \dfrac{i_L}{d}$ ELECTRONS/SEC.
CONDUCTION BAND LIFETIME = τ_ℓ = $\dfrac{n_{oL}}{f_L}$	$= 1.5 \times 10^{17} \dfrac{i_L}{LP}$ SECONDS
TRAP DENSITY $= n_b$ = $\dfrac{e}{kT} n_o \dfrac{\tau_o}{\tau_\ell}$	$= 38.9 \dfrac{n_o \tau_o}{\tau_\ell}$ TRAPS/cm³/VOLT
TRAP DEPTH $= E_b$ = $\dfrac{kT}{e} \ln \dfrac{N_c}{n_o}$	$= 0.059 (19 - \log n_o)$ VOLTS BELOW CONDUCTION BAND

Fig. 6—Trap distribution — CdS powder 1424A.

fall from the conduction band to the recombination centers, but the supply of the conduction electrons will be replenished by electrons boiling out of the traps at a rate determined, at any time, by the depth of the highest filled traps (position of the quasi-Fermi level) and by the density of traps at this level.

Fig. 7—Initial photocurrent decay — CdSe sintered.

At any density of conduction electrons, the corresponding lifetime before recombination may be found from a curve of light intensity versus current, since at equilibrium the rate of generation of electrons by the light must equal the rate of recombination. The time constant for decay gives the time required to remove enough electrons from traps to drop the quasi-Fermi level by kT, since this change corre-

sponds to a reduction of the density of conduction electrons by a factor $1/e$. This corresponds to recombination of the number of electrons filling the traps contained in an energy slice kT wide at the position of the quasi-Fermi limit. Hence, from the ratio of decay time constant to lifetime, we may compute the trap density at the quasi-Fermi level. The depth of the quasi-Fermi level may be computed from the density of electrons in the conduction band.

Fig. 8—Photocurrent decay — CdSe sintered.

This analysis assumes that trapped and conduction electrons are in thermal equilibrium during the decay. This assumption is supported by decay-time measurements made at several points along the decay curve. It has been found that the half-value decay times taken from the complete decay curve are equal to the values found by allowing the

system to reach equilibrium at lower intensities and measuring initial decay rates at these points.

In the tests described here, the light is totally absorbed in a thickness, d, of photoconductor which is small compared with the other gap dimensions. Hence, although the photoconductor sample is relatively thick and irregular, the active region is a rectangular block having length, l, equal to the gap separation, width, w, equal to the length of the electrodes forming the gap, and thickness, d.

Fig. 9—Trap distribution — CdSe sintered.

The formulas for computation of trap distributions are summarized in Table IV.

Representative curves for photocurrent decay and trap distribution curves for a CdS powder sample and a CdSe sintered layer are shown in Figures 4-9. It is noted that the trap densities decrease in a nearly exponential manner with increasing depth below the conduction band. This characteristic is generally observed for these trap distributions.

The values for trap densities depend on the sample thickness in which the light is totally absorbed. It has not been feasible to measure

this with any accuracy. On the basis of various absorption measurements the thickness, d, has been assumed to be 50 microns for CdS and 1 micron for CdSe. The value for CdS is probably too large. Changes in the assumed value of d will result in shifting the trap distribution curves along the ordinate axis without affecting their shapes.

Rose and Lampert[10] suggest that the most probable explanation for the very large M-factor values reported here lies in the assumption that there exists a high density of recombination states in the vicinity of the thermal Fermi level. These states are not in thermal equilibrium with the free carriers. Their occupancy is determined by the kinetics of the recombination process. If their capture cross section for holes is large compared with their capture cross section for electrons (a reasonable assumption for negatively charged states), they will be predominantly occupied by holes. Accordingly, they will not be observed by the present method of measuring trap densities. They will, however, shift the advent of space-charge-limited currents to higher applied voltages and thus allow higher values of gain and, correspondingly, of M values.

---

[10] A. Rose and M. Lambert, *loc. cit.*

MAGNETICS FOR COMPUTERS —
A SURVEY OF THE STATE OF THE ART*†

By

JAN A. RAJCHMAN

RCA Laboratories,
Princeton, N. J.

Summary—The present-day applications of magnetics to random-access memories and logic switching are surveyed and appraised. The memory discussion includes core arrays, external addressing, apertured plates, thin films, and twistors. The survey of switching includes combinatorial switches, shift registers, current steering, voltage drives, transistor coupling, parametron, transfluxors, transfluxor current-steering, and multi-apertured transfluxors.

INTRODUCTION

IN THE last decade the advent of fast-switching square-loop magnetic materials brought about many far reaching inventions in the field of computers and has made of magnetics a major art. New materials, devices, circuits and systems are still being discovered and invented at an unabated rate.

Cores with hysteretic properties have an infinite number of possible remanent states and thus offer the natural possibility of storage of information by means of the value of flux at remanence. Cores which have a square hysteresis loop have, in addition, thresholds of magneto-motive force required to produce irreversible flux changes and offer thus the possibility of artifices for obtaining effective switching. In most applications storing and switching capabilities are exploited simultaneously. For the convenience of this survey, applications based primarily on storage properties, i.e., memories, will be considered separately from applications based primarily on switching properties.

Core Memories

Principle

Arrays of individual toroidal ferrite cores operated in current coincidence are the most widely used type of high-speed memory in present-day computers.[1-3]

Consider an array of cores arranged in rows and columns. Each

Fig. 1—Core memory.

one is linked by three windings. Two of the windings are connected in series by rows and columns and the third is a common series winding. "Single-turn" windings consisting of straight conductors are generally used (Figure 1). All cores are magnetized to one or the other of the two remanent inductions represented by the points P and N on the idealized hysteresis loop of the figure. In this state, the array is holding information previously stored. In order to write in information, i.e., to set any given core to the desired remanence, P or N, pulses of current in the proper direction are simultaneously applied to the

---

[1] J. W. Forrester, "Digital Information Storage in Three Dimensions Using Magnetic Cores," *Jour. Appl. Phys.*, Vol. 22, p. 44, January, 1951.

[2] J. A. Rajchman, "Static Magnetic Matrix Memory and Switching Circuits," *RCA Review*, Vol. XIII, p. 183, June, 1952.

[3] J. A. Rajchman, "A Myriabit Magnetic Core Matrix Memory," *Proc. I.R.E.*, Vol. 41, p. 1407, October, 1953.

row and column of the core. The amplitude of these currents is so adjusted that their additive effects produce sufficient magnetomotive force to exceed the knee of the loop for the selected core at the intersection of the lines, but insufficient to exceed the knee for the other cores in the selected lines where they act singly. Consequently, the selected core will be magnetized in the desired direction while all other cores will remain unaffected.

For read-out, consider the voltage induced in the common winding when write-in currents of a standard direction, e.g., toward P, are sent through the lines of the interrogated core. As flux can change in that core only, the existence of an induced voltage indicates that the core was switched from N to P, while absence of voltage indicates that it was originally at P. The state of the core is ascertained by an interrogation which may change it, a so-called "destructive" read-out mode. To retain information in the memory the interrogated core is restored to its initial state in a second "rewrite" cycle.

The discrimination due to the two-to-one ratio in excitation of the selected core to that of the most-energized unselected core can be maintained as the selection is extended from two to three dimensions. This is utilized in most modern memories to obtain simultaneous access to a number of cores, as follows. If all the cores of a plane array are subjected to a minus one-half current, where one unit of current is the value nominally required to switch-over a core, the application of plus one-half currents to a row and a column will switch no core in the array, as no core is subjected to an absolute excitation greater than one-half. Therefore, if many planes have identical row and column excitations of plus one-half, only on those planes in which the inhibiting minus one-half excitation is absent will the selected core switch-over. There usually is a fourth "inhibit" or "digit write" winding in addition to the read-out and two selecting windings for carrying the inhibiting write-in current (not shown on figure).

Disturb Voltages

Departure from ideal rectangularity of the hysteresis loop of available materials was an important limitation in the development of the core memories. Only moderate rectangularity is sufficient to prevent cumulative demagnetization due to repetitive half demagnetizing steps, so that loss of stored information was never a serious problem. On the other hand, the voltages induced in the read-out winding by the half-excited cores on the selected lines did present a problem, as these "disturb" voltages are additive and may mask out the desired read-out signal, particularly in large arrays. Direct pick-up from the exciting windings adds also to the disturbs. Common practice is to use a

reading winding in which the direction of linkages changes at every core on every line, as in a "checkerboard" pattern, so that the disturb voltages tend to cancel each other. The cancellation depends on the uniformity of core properties. Great efforts have been devoted to methods of fabrication and of testing of cores. Very good results have been obtained in recent years. However, cancellation is in general not a sufficient remedy to the problem of the disturb voltages even with the most uniform cores.

The reversible flux change due to partial excitation of a given polarity is in general different[*] for the two states of remanence of the core so that the degree of cancellation depends on the pattern of remanent states established by the particular stored information. This difficulty can be alleviated by using, after writing, a routine to "post-disturb" the core with a demagnetizing half excitation, a procedure which tends to equalize the reversible flux changes for the two remanent states during subsequent half excitations and thereby improve the cancellation of disturb voltages.[4]

The most widely used artifice to discriminate against the disturb voltages consists of time strobing the output voltage, and is based on the fact that the unwanted reversible flux changes on the half-excited unselected cores occur much faster than the wanted irreversible flux change on the selected core. Another method consists of integrating the read-out signal, preferably over a reading cycle with two periods, to effectively discriminate between reversible and irreversible flux changes.[3] Still other expedients have been proposed. For example, the row and column can be energized one after the other so that only the half-excited cores on the column contribute to the disturb read-out voltage.[5] This expedient favors the use of a rectangular array with long rows and short columns. From all methods to deal with the disturb problem, strobing is the one most frequently used, as it is very effective and does not lengthen the time required for access to the memory.

The problem of the disturb voltages was sufficiently under control several years ago so that arrays of $16 \times 16 = 256$ cores were being operated in computers and arrays of 10,000 cores were being developed in the laboratory.[3] The rectangularity and uniformity of present-day ferrite cores is such that arrays of $64 \times 64 = 4,096$ cores are in common use and arrays of $128 \times 128 = 16,384$ are appearing. In still

---

[*] The difference between these two flux changes is commonly referred to as "delta-flux."

[4] W. N. Papian, "A Coincident-Current Magnetic Memory Cell for the Storage of Digital Information," *Proc. I.R.E.*, Vol. 40, p. 475, April, 1952.

[5] M. A. Alexander, M. Rosenberg and R. Stuart Williams, "Ferrite Core Memory is Fast and Reliable," *Electronics*, Vol. 29, p. 158, February, 1956.

larger arrays, e.g., $256 \times 256 = 65,536$ cores, the problem of disturb voltages is solved by the simple expedient of splitting the read-out winding. A particularly advantageous way to split the winding is by grouping together square subdivisions of the plane along diagonals, as then there are disturb contributions only from the cores on the selected lines contained within the subsquare including the selected core.[6,7]

Performance

The mechanism of switching of square-loop cores is primarily through domain wall motions, at least while it occurs in times longer than 10^{-7} second. The switching time, τ, is related to the magnetizing force, H, by the relation

$$\tau (H - H_c) = S_w, \tag{1}$$

where H_c is the coercive field, and S_w a constant, the switching constant.[8]

In the current-coincidence mode of operation, the total magnetizing force due to the addition of row and column current drives cannot exceed $2H_c$, since the magnetizing force due to the row or column drive alone cannot exceed H_c. In practice it is generally found that an optimum discrimination between full and half excitations is obtained when $H \approx 1.5H_c$. It is seen, therefore, that the switching time is proportional to the switching constant and inversely proportional to the coercive force.

The commonly used cores are made of manganese magnesium ferrite. By varying the composition and heat treatment it is possible to obtain a fairly wide range of coercivity. However, it was not found possible to exert much influence on the switching constant, S_w, which has about the same value of 10^{-6} oersted-second for a wide variety of ferrite materials (as well as metals). Consequently, the present practice consists in making so-called fast materials with relatively high coercive force, and so-called slow materials with relatively low coercive force, but among the various compromises between speed and required driving current the essential figure of merit, S_w, remains constant. This is also true for cores made from thin metal ribbon, as used in early memories.

[6] R. L. Best, "Memory Units in the Lincoln TX-2," *Proc. Western Joint Computer Conference*, p. 160, February, 1957.

[7] J. Saltz and C. S. Warren, U.S. Patent 2,691,156.

[8] N. Menyuk and J. G. Goodenough, "Magnetic Materials for Digital Computer Components, I — A Theory of Flux Reversal in Polycrystalline Ferromagnetics," *Jour. Appl. Phys.*, Vol. 26, p. 8, January, 1955.

Typical cores are ring-shaped and are .025 inch thick with an outside diameter of .080 inch and an inside diameter of .050 inch. Fast cores of this size switch in about 1.5 microseconds and require 900 milliampere turns, while slow cores switch in 5 microseconds and require about 300 milliampere turns. Smaller cores (.050 inch outside diameter, .030 inch inside diameter) used in earlier experiments[3] are becoming generally available; these require less current by an amount in proportion to their diameter.

The access time is usually defined as the minimum time required between consecutive entries — write-ins or read-outs. This time is made up of at least two switch-over times (read-out and rewrite). The time required for decoding the address from its coded form, which is usually binary, into the form appropriate to select the particular array lines is included in the access time in most systems. In addition, reasonable time must be allowed for electronic-circuit "red-tape", i.e., recovery times, tolerances in timing, etc. Memory access times of 5 to 10 microseconds are typical for core switching times of 1.5 microseconds.

Memory Drives

In some of the early core memories, the core arrays were driven directly by tubes. Two tubes and two one-turn windings were necessary at each line to provide the two polarities of drive. This required a large number of tubes, each capable of carrying a current of about ½ ampere, rendering the associated electronic circuitry often more onerous than the magnetic part of the memory.

Many of the later and present memories are driven by magnetic switches.[2-3] Consider an array of switch cores linked by row u and column v windings (Figure 2). All cores are biased by a direct current. The excitation of u or v line alone is just sufficient to cancel the bias and therefore does not switch any core. The simultaneous excitation of u and v switches the selected core. At the termination of the u and v drives the d-c bias restores the selected core to its initial state. Each core of the $u \times v$ array is provided with an output winding which is coupled to row x of the memory array. Another $u \times v$ array is used to energize column y of the memory. The switch accomplishes several functions: (1) it provides part of the address decoding, (2) it provides both polarities of output with single polarity inputs, thus simplifying the driving problem as well as the winding of the memory array, (3) it allows convenient means for impedance transformation as it is relatively easy to provide several turns on the switch cores which are larger and less numerous than the memory cores. This permits better matching to the impedance of the tube or transistor drivers.

Considerable over-all economy results from the use of magnetic switch drives particularly when tube drivers are used and the memory array is large. There are other modes of operating the switch which do not require a d-c bias. For example, in the switch array the selected line v can be driven in the switching direction while all lines u except the selected one can be driven in the inhibiting direction.[9] Alternatively, by the proper connection of outputs in shunted groups it is possible to operate the switch in a so-called set-a-line mode in which

Fig. 2—Current-coincidence biased magnetic switch.

the selected line u is driving and the selected line v is inhibiting.[10] It is also possible to use combinatorial magnetic switches.[2] However, the d-c bias switch is the most widely used.

With the advent of relatively high-current transistors, it became practical to drive the memory directly by these devices. Various types of circuits are in use. In some cases the transistor drives the memory lines directly and provides both polarities of current.[6] In others, a transformer couples the memory line to two driving transistors.[11] In

[9] R. E. Merwin, "The IBM 705 EDPM Memory System," *Trans. I.R.E. PGEC*, p. 219, December, 1956.

[10] J. A. Rajchman, "Ferrite Apertured Plate for Random Access Memory," *Proc. I.R.E.*, Vol. 45, p. 325, March, 1957.

[11] R. E. McMahon, "Transistorized Core Memory," *Trans. I.R.E. PGI*, p. 157, June, 1957.

still others, the transistor "steers" the current from a tube-drive source.[12] There was gradual progress in recent years toward the lowering of the current required to switch cores and at the same time toward increasing the current-carrying capabilities of transistors. These trends favor the direct drive of the memory line by single transistors. For very large arrays, magnetic switches (tube or transistor driven) are still preferable.

Memories with External Word Selection

The storing element of the memory—an individual core as just described or another element such as a hole in a ferrite plate, a patch of evaporated film, or a twistor[13,14] is its own selecting switch when the selection is by current coincidence. Considerable advantage results in certain cases when external switching means are used. Such memories are variously known as "switch-driven," "word-organized," or having "end-on" selection, or having "linear" selection.

Consider an array of cores with row and column windings (as in Figure 1, ignoring the read-out winding). Let each row correspond to a word and each column to bits of the words. Assume that all cores are in states P or N and are storing information previously written in. To read-out a given word, the desired row is energized with an amplitude large enough to switch the cores of the row to one of the two states, e.g., N. Voltages are induced simultaneously on all the columns wherein the cores of the row being interrogated were previously in state P, and only on these columns. These are the read-out signals. To write or rewrite information, the current in the selected row is reversed and reduced to an amplitude insufficient to produce switching by itself. Simultaneously "digit" write currents are applied to the particular column windings corresponding to the cores which it is desired to magnetize to P. The amplitude of these currents is sufficient to produce switch-over on the selected row, but insufficient to alter the state of the cores in the other rows. In other words, both row and column drives are half the required total drive.

In this mode of operation, the read-out step is truly by external word addressing and involves no current coincidence. This has two important implications. In the first place, the read-out signals are free from any disturb signals from cores of other words, as these cores

---

[12] C. S. Warren, W. G. Rumble and W. A. Helbig, "Transistorized Memory Monitors Earth Satellite," *Electronics*, Vol. 31, p. 66, January, 1958.

[13] A. H. Bobeck, "A New Storage Element Suitable for Large-Sized Memory Arrays — The Twistor," *Bell Sys. Tech. Jour.*, Vol. 36, p. 1319, November, 1957.

[14] A. H. Bobeck, "Twistors," Technical Discussion on Solid State Circuits, University of Arizona, May 2, 1958.

are not energized at all. The only undesired signals on the read-out column are due to "elastic" flux excursions of the cores on the selected row in state N being driven further into saturation. The wanted flux change from state P toward the saturated state N is always easily distinguishable from these elastic flux excursions. Consequently, considerable deviation from ideal rectangularity of the hysteresis loop is tolerable and does not impose any limit on the permissible size of the array as is the case with current coincidence operation. In the second place, the read-out current is not limited in amplitude by the selection mode to produce a magneto-motive force equal to 1.5 or 2 H_c, but can be as large as desired, many times larger than the coercive force, H_c. Consequently, it is possible in principle to make the switching time during read-out arbitrarily short simply by making the selecting current sufficiently large (see Relation (1)). The speed is limited by the practical difficulty of obtaining an adequate external switch and by the heating of the storing element should it be switched at the high repetition rate possible with the short switch-over time. Operational memories with read-out switch times of about 0.5 microsecond[6] as well as advanced experiments with times as short as 60 millimicro-seconds have been reported.[15]

The write-in step depends on current coincidence so that the whole addressing is not entirely by external means. The currents, and therefore the speed of writing, are limited. The write step can be shortened by increasing the ratio of switching current on the selected core to the maximum on unselected cores from 2:1 to 3:1. This is done by using $\frac{2}{3}$ of the nominal drive on the selected row and $\frac{1}{3}$ of the drive on all columns, the polarity of the column drive being plus or minus depending on the polarity to which it is desired to magnetize the cores.[2]

The advantages of the word-addressed memory system are obtained at the expense of an external driving means. This external means can consist of transistors, one or two for each word. The use of a single transistor is particularly suitable to obtain very large current for read-out and relatively small reverse current for write-in, yielding typical read-out times of less than a microsecond and write-in times of several microseconds.[6] The driving transistors are themselves driven by diode or transistor decoding switches. Alternatively, the external driving mean can consist of a magnetic switch. In either case the driving means is greatly simplified by the use of two cores per bit instead of one,[6,10] as the voltage induced on the selected row

---
[15] V. L. Newhouse, "The Utilization of Domain Wall Viscosity in Data Handling Devices," *Proc. Western Joint Computer Conference*, p. 73, February, 1957.

can be made independent of the stored information, i.e., the same for every access. The two cores store one bit of information by being either in states PN or states NP. On read-out, a large current is applied to the selected row and brings all cores on the row to state N. This induces a voltage on one or the other column winding of the pair belonging to each digit. By connecting the two column windings in series opposition a positive or negative read-out voltage is obtained

Fig. 3—Switch-driven magnetic memory.

depending on the stored bit. This arrangement automatically neutralizes the effect of the elastic excursions of flux of the cores driven further into saturation. To write-in or rewrite, the current on the selected row is reversed and simultaneously positive or negative currents are sent through the series connected column pair depending on the nature of the bit to be written in.

Word addressing by a magnetic switch has a number of advantages. Figure 3 shows a d-c biased core switch, each core driving a

word of the memory. Figure 4 shows a switch core and its memory
load in the case of a word with a single bit made up of three identical
cores, the switch core and the two memory cores. Assume the resist-
ance of the coupling electric loop to be negligible. Let us assume that
initially the switch core is in state N and the memory cores are either
in states NP or PN. When the selecting currents on the rows and
columns of the switch overpower the d-c bias and switch the switch
core from N to P, whichever memory core was in state P is switched
to state N. This occurs because the total amount of flux linked by a
zero resistance loop must remain constant.[10] The identity of the
memory core being switched determines the polarity of the read-out

Fig. 4—Flux-limited switch drive.

signal on the digit winding. When the selecting currents driving
the switch terminate, the d-c bias restores the switch core from P to
N. This tends to switch both memory cores to P. If there were no
current in the digit winding at that time, the two identical memory
cores would be switched half way, absorbing equally the flux change
of the switch core. A subsequent read-out, due to a new switching of
the switch core from N to P, would switch the two memory cores
equally, and consequently no read-out voltage would be obtained.
However, if during the return of the switch core from P to N, a
current is sent through the digit winding which favors one core and
hinders the other, the flux change in one core will be greater than that

in the other. Consequently, on a subsequent read-out, a voltage is obtained as both memory cores are brought to the standard N condition. The read-out voltage depends on the amplitude and polarity of the write-in current in accordance with a typical characteristic curve, as shown in Figure 4.

The system described above depends on transferring a limited amount of flux from the switch core to the memory cores. This principle has several interesting consequences. The read-out step can be arbitrarily fast as in other types of externally addressed memories. The write-in step is still dependent on current coincidence selection but the two disadvantages of that selection system—limitation in speed and small tolerances in the amplitude of drive—are largely overcome. The pair of memory cores really stores analog information, i.e., gives a read-out output voltage which is function of the digit write amplitude according to the characteristic of Figure 4. For digital storage, only the polarity of the output is ascertained, so that the digit write current can have any value between a minimum which insures no false indications due to inequalities between the cores or to noise, and a maximum which insures that the state of storage of non-selected cores is not altered. The ratio of maximum to minimum permissible amplitude of the write digit pulse is 3 to 1 when the durations of that pulse and the address-selecting pulses are of the length dictated by the conventional current-coincident operation, i.e., typically 1 to 2 microseconds for the faster type ferrites. Actually, it is possible to use very intense address drive producing very fast switch-overs (tens of millimicroseconds) not only for the read-out step, but also for the write-in, as some unbalancing of flux distribution between the two memory cores will occur in principle for any value of digit write current. However, it is preferable to use the greatest permissible value of write current in order to obtain as large a read-out signal as possible. For very short pulses, this maximum value is several time greater than it is for pulses of duration compatible with conventional current operation as was revealed by a recent study.[15] Therefore, the whole read–write cycle can be accomplished in a very short time (tens of millimicroseconds). Unfortunately, corresponding high repetition rates would lead to excessive heat generation in the cores, a factor which gives the ultimate limit to the rate of information input and output flow. This factor limits the repetition rate to about one megacycle with presently available materials and reasonable cooling techniques.

Ferrite Apertured Plate Memories

The arrays of individual cores have become the classical solution

for the selective-access high-speed memory chiefly because they have proven to be much more reliable than previously used systems. Memories with hundreds of thousands of individual cores and some with millions are in use. For these and still larger capacities it appeared some years ago that the technique of individual cores becomes uneconomical. Considerable research to uncover other magetic elements was undertaken. One of these developments, the ferrite apertured plate, has reached an advanced state of development.[10]

A plate with a regular array of holes is molded from square-hysteresis-loop ferrite material. The direction of remanent magnetization around each aperture stores one bit of information. There is no interference between magnetizations around adjacent apertures because, for a given current through an aperture, the magnetizing force diminishes gradually with radial distance and, at a well-defined distance chosen to be less than half the width of the leg between adjacent apertures, becomes smaller than the threshold of switchover. The plate thus constitutes a unit conveniently fabricated as a whole in a single step and which is equivalent to a number of cores. Furthermore, the ferrite being an insulator it is possible to "print" windings directly on the plate and thereby eliminate the need for some of the manual threading. This is particularly suitable for making a winding which links all the holes in series.

The prototype which has undergone most development has an array of $16 \times 16 = 256$ holes. The holes are .025 inch in diameter and are spaced .050 inch center-to-center. The plate is less than an inch square (Figure 5). Squareness of the hysteresis loop is comparable to that of individual cores made of the same material. Very uniform properties from hole to hole and plate to plate have been obtained. About 300 milliampere-turns are required to switch over the magnetization around an aperture. The switching time is about 1.5 microseconds.

The plates are most conveniently stacked with all the holes in register (Figure 6). The whole stack can be threaded in a single operation which does not need to be repeated separately for each plane as is customary for memory core planes. This also avoids the making of very many soldered connections. Stacks composed of 1, 4, 16, or 64 plates constitute convenient modules with which memories of various capacities can be built. Plate memories can be used in current-coincident operation or external word selection.

For current-coincidence operation the holes of the stack of plates are threaded back and forth by row X and column Y windings. With this wiring pattern the direction of coupling of the selecting lines

is the same as that of the printed winding and, therefore, during readout the disturb voltages would add up rather than tend to cancel each other. The disturb signals can be cancelled by using one hole in each of a pair of plates for every digit with digit windings connected in series opposition. Each bit is stored by the pattern PN or NP in a pair of corresponding holes. The polarity of the net signal is indicative of the stored bit. For writing or rewriting, an inhibiting current is sent through the common winding of one or the other plate. Another method of dealing with the disturb signals recently devised[16] uses only one hole per bit. Four plates are used in the smallest module. The plates are arranged in a square and are con-

Fig. 5—Ferrite memory plates.

nected in series diagonally in pairs. The two diagonal pairs are connected in series opposition and the mid-point is grounded. A stack of quartets of plates is wired as a whole by X and Y selecting windings. In the reading step, the polarity of the difference voltage between the two non-grounded terminals of the diagonally connected plates is sensed by a difference amplifier and is indicative of the identity of the stored bit. The disturb signals tend to cancel. For writing or rewriting, the inhibit signal is sent into both branches of the common winding, in one polarity in one and the other polarity in the other. Extremely good discrimination was obtained with this

[16] W. G. Rumble and C. S. Warren, "Coincident Current Applications of Ferrite Apertured Plates," *I.R.E. Wescon Convention Record*, Part 4, p. 62, 1958.

arrangement. This is due in part to the high uniformity of the plates and in part due to the lack of any appreciable air coupling. Planes of that type including as many as 16 plates yield satisfactory discrimination ratios.

The apertured plates are particularly suitable for external word addressing as the switch can be made of plates identical to those used for the memory. The memory stack and the switch stack can be wired as a whole with a bundle of wires going through all the holes.

Fig. 6—Stack of memory plates.

The switch stack is wired in addition with X and Y selecting windings (Figure 6). The switch can be operated in a number of ways such as set-a-line system,[10,17] and the d-c bias system.[10,18] The latter seems to be the most advantageous. Bias current can be conveniently supplied through the printed winding. The memories with a d-c bias switch and a pair of plates for each digit were made according to

---

[17] W. J. Haneman, J. Lehman and C. S. Warren, "Apertured Plate Memory: Operation and Analysis," *I.R.E. National Convention Record*, Part 4, Vol. 6, p. 254, March, 1958.

[18] M. M. Kaufman and W. L. Newhouse, "Operating Range of a Memory Using Two Ferrite Plate Apertures per Bit," *Jour. Appl. Phys.*, Vol. 29, p. 487, March, 1958.

the system explained previously in connection with core memories. In this case, the operation is not strictly flux limited, since the amount of flux available from a plate hole increases monotonically with the drive. It turns out, however, that very fast operation and larger tolerances in current are possible with the plates than with cores.[18] Here again the limit in speed is due to power dissipation limitations rather than due to the system of selection.

Apertured plates are more economical to fabricate and test than are the corresponding number of individual cores. Also their assembly and wiring into memories is considerably simpler. Furthermore, a smaller driving energy is necessary as it is possible to use a much smaller amount of material to store one bit than is necessary with individual cores which become unmanageable when very small. The present state of development of the plates is very advanced. Fabrication techniques have been mastered. Experimental memories of the switch-driven and current-coincidence types have undergone exacting tests. It appears that the plates might offer an order of magnitude of improvement as compared with individual core memories.

Thin-Film Memories

A recent development of great significance has been the utilization of evaporated thin metallic films of magnetic materials for the elements of memories.[19] This development is based on the expectation that this technique can provide at once a very simple solution to the problems of fabricating very large arrays by a single step and of obtaining high speed of operation.[20,21]

Patches of magnetic metal are deposited on a heated glass substrate in the presence of a d-c magnetic field in the plane of the substrate. The resulting film develops a preferred magnetic axis in the direction of the orienting field. Films made of 80-20 permalloy about 1,000 to 2,000 Å thick are evaporated in circular patches of about 5 to 10 millimeters in diameter to form a regular array of dots.

The material exhibits a very square hysteris loop. The demagnetizing field due to free poles at the edges of the film, which tend to shear the hysteris loop, have a negligible effect as long as the ratio of length-to-thickness of the film is very large. For 2,000 Å

[19] A. V. Pohm and S. V. Rubens, "A Compact Coincident Current Memory," *Proc. Eastern Joint Computer Conference*, p. 120, December, 1956.

[20] M. S. Blois, Jr., "Preparation of Thin Films and Their Properties," *Jour. Appl. Phys.*, Vol. 26, p. 975, August, 1955.

[21] D. O. Smith, "Static and Dynamic Behaviour of Thin Permalloy Films," *Jour. Appl. Phys.*, Vol. 29, p. 264, March, 1958.

films, spots of 3 to 4 millimeters diameter have been reported to be necessary to prevent serious shearing of the loop.[19] The switching mechanism which has been extensively studied theoretically and experimentally is believed to be due to simple rotation of the total magnetization and involve no domain wall motion.[20,21] The circular shape of the storing elements is chosen to eliminate anisotropy effects as the magnetization undergoes rotation during switching. The switching constant, S_w, is found experimentally to be about .1 oersted-microsecond, or about 10 times smaller than for most of the conventional ferrites. There are theoretical reasons to believe that a still smaller value could be obtained. Furthermore, switching is enhanced by the presence of a relatively small magnetic field in the plane of the film and perpendicular to the easy direction of magnetization.

The array can be driven by thin flat conductors; along the surface of these conductors, fairly uniform magnetic fields are produced.[19] The magnetic elements are placed in close proximity to the conductors. At the element position, the row, column, and inhibit conductors, are made to run parallel to each other. As a result, the effects of the driving currents are algebraically additive and allow operation in the conventional current-coincident mode. The windings can be made by conventional photo-etching printed-circuit techniques. To minimize the drive-current requirements and the inductance of the drive lines, windings are placed both on top and on the bottom of the supporting glass sheet which is made as thin as possible. The drive lines in the region of the elements are slit to prevent eddy currents due to the flux return which links across these lines and the neighboring space. These lines appear as transmission lines with a characteristic impedance of tens of ohms and a propagation velocity of about 2×10^8 centimeters per second. Line drive currents of about 400 milliamperes are required to produce the necessary fields of 1 oersted. The enhancing transverse field is obtained by inclining at a small angle the direction of the winding at the location of the elements with respect to the easy direction of magnetization. The sense winding is made to link the elements in checkerboard fashion so as to neutralize the disturb voltages. There are delta-flux effects; however, it is reported that theoretically these can be eliminated by proper geometry of the sense winding.[19]

Typical experimental results[19] from a 4-millimeter dot in an actual array are: switching in .5 microsecond and signal of 4 millivolts in a sense winding having a characteristic impedance of 20 ohms.

Experiments with single dots have tended to confirm the theoretical prediction of very short switching times. Switching in 10 millimicroseconds by fields of only 2 oersteds have been reported,[21] indicating a

very low value of the switching coefficient, S_w. Fast switching with a moderate drive and the fact that corresponding high repetition rates are not prevented by adverse heating because of the favorable surface-to-volume ratio of thin films, has offered the possibility of very fast memories. However, there are some practical difficulties in exploiting this potentiality. Because of the air return flux path the diameter of the dots cannot be too small (typically 3 millimeters) nor can they be too closely spaced (6 millimeters center-to-center). The resulting physical size of the array tends to make the propagation time of the driving signal longer than the switchover time of the element.

The energy required to switch-over the film can be very small, as the switching constant is low and the volume of the material is small. With proper impedance matching, this can lead to small power required from the driving circuit. On the other hand, there are serious practical problems in sensing the resulting very small read-out signal which tends to be obscured by large signals due to direct pick-up from the driving lines. These extraneous signals can be neutralized by a very careful balancing provided the geometry of the windings is sufficiently precise.

The answer to questions relating to uniformity of the elements, number of bits permissible in a plane, reliability, ultimate cost, etc. await further engineering.

The Twistor

A way to utilize a magnetic wire for the elements of a random-access memory has been reported.[13] Consider a magnetic wire or ribbon wrapped around an insulated nonmagnetic wire (e.g., enameled copper) in the form of a helix making an angle of 45° with respect to the axis. Let there be in addition a solenoid around the wire (Figure 7). This constitutes a form of twistor, which can be called "barber-pole" type. The combined effect of a current I_s through the solenoid and of a current I_w through the wire can produce a magnetizing force parallel to the helical wire. The lines of flux will follow the easy path along the helical wire and return from the ends of the wire through the space surrounding the element. The hysteresis loop is very rectangular due to the anisotropy of the wire created by its longitudinal tension. The air return path has no serious shearing effect on the loop as long as the ratio of length-to-diameter of the element (enhanced by the helical path) is sufficiently large to insure that the demagnetizing field is smaller than the coercive force.

Originally, a form of twistor which consisted simply of a twisted magnetic wire was reported.[13] The twist causes lines of strain at 45°

with respect to the axis. The resulting anisotropy produces an easy direction of magnetization either in the direction of maximum tension or maximum compression, depending on the sign of the magneto-striction coefficient. In any case an easy direction along a 45° helix is obtained and permits twistor action by the combination of a current through the magnetic wire itself and through a solenoid wrapped around it. As the operation is dependent on the amount of twist, this seemingly simpler type of twistor is not as convenient to use in large memories as the barber-pole type. In the latter the tension of the magnetic wire is permanently set as that wire is machine-wrapped around its supporting axial copper wire, and elements with uniform properties are easily obtained.

The twistor can be operated in various modes. Helical magnetization can be produced either (1) by the coincidence of a current I_w

Fig. 7—Twistor.

through the wire and a current I_s through the solenoid (mode A), or (2) by the coincidence of current through two solenoids I_{s1} and I_{s2} wrapped around the twistor (mode B), or (3) by either the wire or the solenoid currents alone. Each of the writing-in and read-out steps can be executed by either mode making it possible to have a number of memory systems.[13]

Development of a memory with external word selection for which the twistor is particularly suitable was recently reported.[14] The barber-pole twistor wires correspond to the digits of the word. The words are selected by coils which link the twistors by lines and which produce the axial magnetic fields being essentially one-turn solenoids (Figure 8). Read-out is obtained by sending through the selected word line, a current I_s intense enough to switch-over the twistors which it links. The twistors which were previously magnetized in the

opposite direction, switch-over and produce a read-out signal on the corresponding twistor central wires. These signals are relatively intense, because the magnetic wire wraps around many times the "single turn" central wire. This gain in voltage is a unique property of the twistor. The twistors which were previously magnetized in the direction of the read-out drive undergo small reversible flux charges whose radial component is of the opposite sign from the radial component of the irreversible total flux change. This apparently surprising effect, which gives rise to a negative slope of the effective hysteresis loop at remanance, can be explained by considering that in this region the flux changes by coherent rotation. The write-in is obtained by a current coincidence mode. The effects of

Fig. 8—Twistor memory.

the current I_s in the word select line and of the current I_w in the central digit wires add algebraically. Additive, subtractive (inhibitive) or additive–subtractive combinations are possible as is the case with any type of externally word addressed memory.

In the announced memory being developed[14] the word selecting is by a ferrite-core d-c-biased switch. The barber-pole twistors are made by wrapping on .003-inch insulated copper wire a .001-inch permalloy wire flattened to .0003 inch. These wires are imbedded in parallel lines spaced ten to the inch between two sheets of polystrip plastic. Similar polystrips with imbedded ribbons of copper wire are placed on either side of the twistor sheets to provide the words selecting coils. A memory of 4,000 words requires polystrips 60 feet

long which can be folded into a volume of about ⅓ cubic foot. The current I_s in the selecting coils is of the order of amperes and the current I_w in the central twistor wires of the order of tenths of an ampere. The readout signal is a few millivolts, and the cycle period is about 6 microseconds.

The twistor utilizes anisotropies produced by ingeniously made stresses in magnetic metals. It has opened new fields of application in magnetics. Operating twistor memories of fair capacity are likely to be a reality in the near future and may represent a considerable step forward beyond the conventional individual-ferrite-core memories.

Magnetic Switching

Magnetic elements can be used for switching as well as storage as was mentioned earlier. These applications have been recently surveyed in some detail[22] and are considered only briefly in the present paper.

Cores Directly Driven by Electronic Devices

In high-speed applications, magnetic elements must be driven by electronic devices—tubes or transistors. Perhaps the simplest applications are those in which the cores are driven and sensed directly by electronic devices, and do not drive each other. The memory arrays described are outstanding examples of such directly driven circuits. Other examples are combinatorial switches for the purpose of selecting one, or several outputs from a large group. This is accomplished by using a number of windings, each linking certain cores in series in such a way that energizing selected groups of windings will produce an algebraic sum of magnetomotive forces which exceeds the threshold of switch-over on the selected core only. There is a considerable difference between such switches and the current-coincident memory, which is also combinatorial, in that the non-selected cores in the switch can support arbitrarily large magnetomotive forces tending to drive them in the direction of their existing saturation, whereas this is not permissible in the memory arrays because the state of remanence is unknown by definition. Typical are the following types of combinatorial switches.

A switch to select one out of $n \times m$ outputs by the choice of one input among n, and one input among m, has already been mentioned. It utilizes an array of $n \times m$ cores linked by n rows and m columns and also by a common winding carrying a d-c biasing current.

A variety of so-called decoder and encoder combinatorial switches

[22] J. A. Rajchman, "Magnetic Switching," to be published in *Proc. Western Joint Computer Conference*, May, 1958.

can be made.[2] A typical case would be a switch with eight outputs and three inputs whose purpose it is to select one output for every possible combination of inputs (Figure 9). For simplicity, the cores are represented on the figure by heavy lines and the linking windings by 45° segments of line. The input signals are in pairs, each one linking half the cores. The first input links the cores by juxtaposed halves, the second by interlaced quarters and the third by interlaced eighths. The direction of winding is such, that a current sent in one or the other branch of each input tends to magnetize the cores further into saturation. Consequently the application of the inputs have no effect per se. If, during the presence of the inputs, all cores are energized in a direction tending to reverse their magnetization, only the core which is not inhibited by the input currents will have a net reversing magneto-

INPUTS

		1	2	3	4	5	6	7	8
	1	+	+	+	+	+	+	+	+
	2	+	−	+	−	+	−	+	−
	3	+	+	−	−	+	+	−	−
CORES	4	+	−	−	+	+	−	−	+
	5	+	+	+	+	−	−	−	−
	6	+	−	+	−	−	+	−	+
	7	+	+	−	−	−	−	+	+
	8	+	−	−	+	−	+	+	−

Fig. 9—Pattern of windings for a load-sharing switch.

motive force and therefore will be the only one to switch over. After the core has been switched, it can be restored by energizing a winding which links all cores. Decoding switches of this type have been used as input switches for magnetic core memories and many other applications. The usefulness of the binary decoder switch can be broadened by using output windings that link in series certain cores according to a desired code (Figure 9).[23] Switch-over of the selected core will cause output signals on only those output windings which are coupled to it. Any truth table relating any n inputs to any m outputs can be obtained in this fashion, demonstrating thereby the feasibility of obtaining any desired logical function.

Magnetic switches can be made to utilize the power from several sources to drive a single load. Such "load sharing switches"[24] permit

[23] J. A. Rajchman, "A Survey of Magnetic and Other Solid-State Devices for the Manipulation of Information," *Trans. I.R.E. PGCT*, p. 210, September, 1957.

[24] G. Constantine, Jr., "A Load-Sharing Matrix Switch," *IBM Journal of Research and Development*, Vol. 2, p. 204, July, 1958.

the use of relatively low-powered drivers to furnish a relatively high power, and may be used for example for transistor drive of core memories. A number N, where N is an integral power of 2, sources can be made to drive an equal number of cores in such a way that one selected core receives the total excitation from all the sources while the rest of the $(N-1)$ cores receive zero excitation. Figure 10 illustrates the wiring arrangement for the case of $N = 8$. The signs $+$ and $-$ refer to the direction of linkage through the core of each of the eight input windings linking the cores in series. The currents from the source are assumed to be $+$ or $-$ one unit of current (with unipolar current sources a pair is necessary for each input). It is clear from inspection

Fig. 10—Combinational core switch.

that if the input circuits are excited according to the pattern of polarities corresponding to any one of the eight patterns of winding polarities of the cores, the selected core will have eight units of excitation while other cores will have precisely no excitation. In such a switch the combinations of windings are exploited for load-sharing purposes rather than to gain a combinational advantage in coding or decoding. In general, the inputs to such a switch must be decoded by some other means. It should be mentioned that there are no particularly severe requirements on the nonlinearity of the cores, as the saturation characteristics are useful only to attenuate the effects of whatever inequalities of amplitude there may be between inputs.

Magnetic Shift Register

The classical example in which the switching of one core causes the switching of another is the magnetic shift register.[25] The operation of this device, in which the pattern of settings of a row of cores can be shifted along the row, is well known (Figure 11). The information-carrying quantity is the magnetic flux. Flux is transferred from core to core without any loss because it is possible to make up for the ohmic losses in the coupling circuits by an excess of turns on the winding on the input core with respect to those on the output core. There must be provision for forcing the information to flow in the desired direction. Furthermore, the core to which flux is being transferred must not be loaded by the next core of the row, as this loading would

Fig. 11—Shift register.

prevent its proper setting. To prevent this back flow of information and detrimental loading, diodes are used in the coupling circuits.

The example of the shift register shows that the addition of diodes to magnetic cores provides power gain and permits transfer of information from core to core indefinitely. The gain results from the fact that the switching of a core to one polarity produces no load current and requires, therefore, only the amount of energy necessary to make up the losses in the core, whereas the switching of the core to the other polarity does produce a load current and transfer of energy from the source to the load. The logical input "primes" the core and determines thereby that an output is produced during the "drive" of the core. An unprimed core produces no output during drive.

[25] A. Wang and W. L. Woo, "Static Magnetic Storage and Delay Line," *Jour. Appl. Phys.*, Vol. 21, p. 49, January, 1950.

The decoupling of the loading core during priming can also be achieved by using a resistive coupling instead of a diode; in operation, priming is at a relatively slow rate so as to make the resultant induced voltage (and current) small, and driving is relatively fast so as to make the resultant voltage (and current) high. In this way, the resistance seems to be in the circuit during priming, but seems to disappear during driving. Coupling cores between the switching cores can be used with advantage to realize this artifice.[26,27] Multiple-hole cores can be used to make diodeless and resistanceless shift registers as described later.

The solution to the problem of transferring flux from core to core in the magnetic shift register can be applied to more complicated networks of interconnected cores which can perform a great variety of switching functions. Many such circuits have been built.

Current Steering in Magnetic Circuits

Circuits[28,29] can be made in which the current from a source can be steered through one particular path among many possible paths. Consider the lower circuit of Figure 12. Let all cores be set in the normal direction except one which is set in the abnormal direction. A pulse of current sent through the circuit, and flowing through the common winding in a direction tending to bring all cores to the normal state, will switch over the abnormal core. This will induce a voltage on its output winding which, with the proper choice of polarities, will tend to make the corresponding diode conducting and to cut off all other diodes. As a result, the current will flow entirely through the selected branch.

For the steering to be successful, there must be more turns on the switching winding than there are on the branch windings (not shown in the simplified figure) in order that there be an excess of magneto-motive force producing switch over in spite of the selected branch current which necessarily opposes the effect of the driving current.

The prior setting of the core that selects the conductive branch produces no branch current in itself since the back-to-back diodes block any possible current flow. This setting is akin to priming the circuit. The desired branch current has an amplitude which is exactly the

[26] G. R. Briggs, "A Magnetic Core Gate and its Application in a Stepping Register," *MIT Digital Computer Eng. Note E-475*, October, 1952.

[27] L. A. Russell, "Diodeless Magnetic Core Logical Circuits," *I.R.E. National Convention Record*, March, 1957, Part 4, Vol. 5, p. 106.

[28] J. Karnaugh, "Pulse-Switching Circuits Using Magnetic Cores," *Proc. I.R.E.*, Vol. 43, p. 570, May, 1955.

[29] J. A. Rajchman and H. D. Crane, "Current Steering in Magnetic Circuits," *Trans. I.R.E. PGEC*, p. 21, March, 1957.

amplitude of the drive current source since the branch current is the drive current itself. Therefore it is independent of the core and diode properties.

The usefulness of the principle of current steering can be illustrated by its use in a binary decoder switch (Figure 13). The three pairs of selecting windings that were driven by tubes in the previous example are now driven by current steering. Each tube is replaced by a core and diode. The parallel pairs of selecting windings are connected in series. The inputs to the switch set one core in each pair differently from the other. The drive brings all the input cores to a standard direction of magnetization, thereby switching one core in each pair. This causes the drive current to be steered through one branch of every pair. Con-

Fig. 12—Principle of current steering.

sequently, all output cores but one will be subject to inhibiting currents and the selected one will be subject to a switching current, just as was the case in the tube-driven decoder. That particular core will switch over and produce a signal in its output winding.

Another example of a current-steering circuit is one in which new settings are produced by the steered current itself. It is a "two-phase stepping register" which could also be thought of as a current-steering commutator (Figure 14). The purpose of the circuit is to deliver a given current successively to a number of loads. The steering cores are divided into two groups. A first drive current, steered by a given core, k, of one group flows through a corresponding load and switches

a core, $k + 1$, of the second group. In turn, a second drive current steered by the previously set core, $k + 1$, flows through its corresponding load and sets a new core, $k + 2$, in the first group and so on. The circuit is similar to the conventional magnetic shift register (used as a counter) with the important difference that the advance current itself is steered through the successive loads.

Voltage-Driven Circuits

The combinatorial switches, magnetic shift registers, and commutator switches described thus far were all driven from current sources.

Fig. 13—Current-steering decoder switch.

A current source is a device furnishing a given current without regard to the voltage this current will produce in the load, hence can be considered as a generator with infinite internal impedance. The current-steering principle epitomizes the current-driven-type circuits. Magnetic switches can also be driven by voltage sources, that is, by generators which produce a given voltage without regard to the current produced in the loads and which can therefore be considered to have zero internal impedance.

Some of the well-known voltage-driven magnetic computer circuits depend on the use of the core as a synchronous magnetic amplifier.[30]

---

[30] R. A. Ramey, "Magnetic Amplifier Circuits and Applications," *AIEE Conference Paper*, 1953.

Fig. 14—Current-steering commutator.

Consider the core of Figure 15 with an input and an output winding and with diodes in each circuit as shown. Power is provided to both windings, i.e., a-c sine waves of the same frequency and phase. Square waves or any other symmetrical waves can be used also. The two voltages E_{ac} and E'_{ac} have, at any instant, the relative polarities shown and have amplitudes proportional to the number of turns of their respective windings. The operation of the amplifier is in two steps — reset and gating. During the half-cycle with polarities as indicated in the figure, current can flow only in the input or reset circuit because of the direction of the rectifiers. This causes the magnetic core to proceed from saturation (or reset state) an amount dependent on E_s and E'_{ac}. If E_s is zero the core is switched over by E'_{ac}, if E_s is equal to E'_{ac} there will be no voltage across the cores since E_s and E'_{ac} will cancel so that the

Fig. 15—Principle of voltage-driven circuit.

core will not switch over. When E_s is zero and the core switches over, there is no current in the output winding even though the direction of the diode would permit it because the voltage E_{ac}, which is greater or at least equal to the voltage induced in the output winding, keeps the diode cut off.

The second or gating step is in the next half-cycle. The voltages are the reverse of those shown in the figure. If the core was not switched over in the preceding step $(E_s = E'_{ac})$, then the voltage E_{ac}, which tends to reset the core, appears almost entirely on the load, while the load current simply brings the core further into saturation. If the core was switched over in the preceding step $(E_s = 0)$ then the voltage E_{ac} will appear almost entirely across the output winding with only a small voltage drop across the load, due to the magnetizing current. It is seen, therefore, that if there is no input $(E_s = 0)$ there is practically no output $(E_L = 0)$ and if there is input $(E_s = E'_{ac})$ there is an output $(E_L = E_{ac})$. In the conversion of the intelligence signal from the input to the output, several quantities can change. (1) The voltage E_{ac} can be greater than E'_{ac}; (2) there can be more power in the output than the input because the input needs to provide only enough power to magnetize the core while the output can provide far in excess of that amount; and (3) the output occurs one-half cycle after the input.

Voltage-driven circuits have been investigated quite thoroughly and all-magnetic computers have been built, using unit blocks each including basically voltage-driven single core magnetic amplifiers and a number of diodes (Ferrators).[31] Some of the diodes provide the necessary elements required for obtaining gain and isolation from other units, while others are used for logic switching.

Transistor-Coupled Circuits

In the magnetic logic circuits described above, the power originates from a central source, all elements of the circuit being passive and dissipating power. They require only a very few driving tubes by centralizing all sources of gain and power and, therefore, minimizing the expense and possible trouble. This is in keeping with the philosophy that the complexity of a circuit is measured by the number of tubes, the tubes being the most expensive and the least reliable element of the circuit. With the advent of the transistor this philosophy may no longer be valid, as it becomes practical to contemplate many local elements with real gain interspersed between passive elements.

An example of such a circuit is a magnetic shift register with

---
[31] T. H. Bonn, "The Megacycle Farrator," *Proc. Western Joint Computer Conference*, p. 118, February, 1956.

transistor coupling between cores as shown in Figure 16.[32,33] Actuation of one of the advance windings, e.g., *A*, causes the abnormal core (or cores) to switch from *p* to *n*, as in the previous example of a core-diode shift register. This induces a voltage on the winding connected to the base of the corresponding interstage transistor and thereby makes the transistor conducting. The relatively large collector current of the transistor switches over the next core from *n* to *p* and thereby effectively causes a shift of information. It is possible to improve the operation by a feedback coupling winding (shown dotted on the figure) on the core being reversed by the advance winding, by means of which the collector current of the transistor not only tends to switch over the next core, but also helps in turning over the core being actuated by the advance winding. This produces a "snap-action" and results in a greatly reduced value of required advance current, which now needs only to "trigger" the core.

The Parametron

Switching of digital information can be performed by magnetic amplifiers of the conventional type which are energized by a carrier of frequency higher than the maximum repetition rate of the pulses representing the digits. A simple type of carrier-energized amplifier which can be used as a basic computer element is the recently announced parametron.[34] It requires no diodes.

A parametron element consists of two cores made of nonlinear material each provided with an input and output winding, the input windings being in series and energized by a current of frequency $2f$, the output windings being connected in series opposition and connected to a condenser which nearly resonates the output circuit to the frequency f (Figure 17). With this arrangement a steady-state oscillation of frequency f is induced in the output circuit having the usefull property that this oscillation can be in either of two stable phases opposite to each other. The identity of the phase is used to represent the two digits, 0 and 1.

When steadily energized, the parametron represents a binary storage element. To store the desired digit, i.e., to establish the desired phase of oscillation, the energization of frequency $2f$ is momentarily

[32] S. Guterman and W. M. Cavey, "A Transistor-Magnetic Core Circuit; a New Device Applied to Digital Computing Techniques," *I.R.E. National Convention Record*, March, 1955, Part 4, Vol. 3, p. 84.

[33] H. Kihn and W. E. Barnette, "The Megacoder — A High-Speed Large-Capacity Microminiature Decoder for Selective Communication," *AIEE Fall General Meeting*, Chicago, Ill., October 7-11, 1957.

[34] Eidi Goto, "On the Application of Parametrically Excited Non-Linear Resonators," *Denki Tsushin Gakkai-shi*, October, 1955.

Fig. 16—Transistor-coupled shift register.

weakened or removed and a relatively weak signal of frequency f and of the desired phase is coupled to the output circuit. Then the energization is restored the full steady-state amplitude of the signal of the desired phase is established.

As seen from the operation as a storage element, the parametron can act as an amplifier, i.e., increase the amplitude of a signal of the desired phase. The device can be used also as a basic logic gate. Consider the input of a parametron to be a linear combination of the output of two parametrons x and y and of a reference parametron of fixed voltage amplitude. The driven parametron will assume the phase of the majority of the inputs and therefore will be an "AND" gate for the signals x and y if the reference phase corresponds to "zero" and an "OR" gate if the reference phase corresponds to a "one." Logic gates for more than two signals are possible also.

In logic circuits, parametrons are coupled through linear impedances in a manner designed to provide the desired logic function. The parametrons are in 3 groups, I, II, III, each group being supplied by

Fig. 17—Parametron.

a separate source of energization at the frequency $2f$. The sources are switched-on in a sequence as follows: I, I II, II, II III, III, III I, I, etc. The states of parametrons of group I influence the states of the parametrons of group II when sources I and II are on and are not influenced by, and do not influence, the parametrons of group III since these are not energized during this time. The newly established states of group II are maintained as energization of group I is removed. A new logical transfer is produced when sources II and III are on, etc. This triple time excitation method insures that there is direction to the information flow.

Computer systems comprising several thousand parametrons have been reported.[35] The exciting frequency was about 2.2 megacycles and the repetition rate of the clock 10 to 30 kilocycles. Only 50 milliwatts of power was required per parametron.

The parametron permits the making of all-magnetic computers. Reliability and economy are the chief advantages, while relatively slow speed is the main disadvantage of this device.

The Transfluxor

Magnetic flux, the information-carrying quantity, can be manipulated by geometrical transfers in branches of cores having a more complicated configuration instead of by electrical linkages between simple ring-shaped cores. This is the underlying principle of the transfluxor.[36,37]

In a two-apertured transfluxor having three legs, the amount of remanent flux in leg 1, which is equal to the negative algebraic sum of fluxes in legs 2 and 3, determines how much transfer of flux between legs 2 and 3 is possible (Figure 18). There can be an indefinitely long back-and-forth exchange of the flux which was initially set in. The transfluxor is blocked when both legs 2 and 3 are saturated in the same direction as one or the other leg denies any of the flux flow. The amount of setting depends on the amount of flux that can be changed in one leg before it is saturated. That same amount will necessarily appear in the other leg. The device can be used to store analog as well as digital information. Only digital applications are considered here.

The nondestructive read-out property of the transfluxor permits making magnetic switches for selecting channels of transmission through which information can be transmitted for as long a time as

[35] Saburo Muroga, "Elementary Principle of the Parametron," *Datamation*, September/October, 1958.
[36] J. A. Rajchman and A. W. Lo, "The Transfluxor — A Magnetic Gate With Stored Variable Setting," *RCA Review*, Vol. XVI, p. 303, June, 1955.
[37] J. A. Rajchman and A. W. Lo, "The Transfluxor," *Proc. I.R.E.*, Vol. 44, p. 321, March, 1956.

desired. A class of such switches or storage systems is obtained by substituting a transfluxor for a core in the combinatorial core switches and memories described previously. As a result, a channel for bilateral transmission of signals is obtained in the selected element instead of a single transient at the instant of selection. Examples illustrating this possibility follow.

Fig. 18—Principle of transfluxor.

Nondestructive Read-Out Random-Access Memory

An array of transfluxors can be linked with one set of column and row windings through the large input apertures, and another set through the small output apertures (Figure 18). Current-coincidence setting of the transfluxors is possible because there is a threshold of current required to produce setting, and similarly, current coincident read-out can be obtained because there is a threshold of current required to obtain read-out. This type of memory is particularly useful when frequent and rapid read-out is necessary and relatively slow write-in can be tolerated.[37]

Cross-Bar Switch

An array of transfluxors can also be utilized for a cross-bar switch (Figure 19). Pulse- or amplitude-modulated signals applied to the column windings linking leg 3 will be transmitted to those row windings linking leg 3 which are coupled through unblocked transfluxors, but not those coupled through a blocked transfluxor. The setting of the transfluxors can be obtained by using the column and row windings linking leg 1.[37]

Fig. 19—Non-destructive memory; cross-bar switch.

Coded Channel Selector

A decoder switch analogous to the core type can be made by using transfluxors. The large aperture is treated as if it were a simple core and is threaded by the input selecting setting windings. As a result, a selected transfluxor can be set and all other blocked. Therefore, a modulated signal fed through the small hole of all transfluxors will be transmitted only to the one that is set.[37]

Channel Commutator

A device for successively opening a number of channels can be made by using a row of transfluxors. Here again the large apertures are regarded as simple cores and are connected as the cores of a shift

register (Figure 20). Current-steering type connections are particularly suitable. After the termination of an advance pulse, one of the transfluxors is set while all others are blocked. Consequently, modulated signals can be transmitted through that channel only, and will continue to be so transmitted until a new advance pulse selects the next channel.[38]

Fig. 20—Channel commutator.

Control of Light Sources

The transfluxor is particularly adaptable to the control of light sources, such as incandescent lamps, gas discharge lamps, or electro-luminescent cells. Let the light source be coupled to the output of a constantly energized transfluxor (Figure 21). Light will be emitted in accordance with the amplitude of the setting, which is zero when the transfluxor is blocked and which can have any desired intermediate value up to a maximum. Possible applications, include indicators to

---

[38] J. A. Rajchman, "Principles of Transfluxor and Core Circuits," to be published in *Proc. International Symposium on Switching Theory*, Cambridge, Mass., April, 1956.

monitor the occurance of pulses in digital systems, small arrays of lights to display numerals or letters, or large arrays to display more complex patterns and pictures.[39]

Transfluxor Current Steering

The above examples show the usefulness of the nondestructive read-out property of the transfluxor. The isolation of the input setting and the output circuit of the device can be exploited for current steering.[29]

Transfluxor steering is achieved by the use of a transfluxor winding linking leg 3 connected in series with a branch load in each of a number of parallel branches. The steering action can be considered to be due to the difference in effective impedances of a blocked and set transfluxor.

Fig. 21—Transfluxor control of light sources.

By setting the two transfluxors to different states, a preferential current flow is obtained. However, diodes are not required since the setting of the transfluxors causes no voltages in the output windings.

Transfluxor steering of current through one branch of a parallel pair can be repeated in many pairs connected in series and is directly applicable to driving a decoding switch. The example of a decoder with $n = 3$ inputs and $2^n = 8$ outputs used to illustrate the tube and core–diode decoders is used again in Figure 22 to illustrate the transfluxor steering decoder. After one transfluxor in each pair is set by the inputs, the DRIVE current is steered through the branch of each pair including the blocked transfluxor and thereby causes a net reversing magnetomotive force on only the single selected output core.

---

[39] J. A. Rajchman, G. R. Briggs, and A. W. Lo, "Transfluxor Controlled Electroluminescent Display Panels," *Proc. I.R.E.*, Vol. 46, p. 1808, November, 1958.

Multi-Apertured Transfluxors

The principle of the geometrical flux transfer can be generalized to cores with more than two apertures and three legs.[36-38, 40-43]

A third aperture added to the two of the transfluxors described above can be used to eliminate any possibility of detrimental overset-ting.[37] Legs 1, 2, 3, and 4 of the transfluxor illustrated in Figure 23 are of equal cross section. The amount of flux that can be set is limited

Fig. 22—Transfluxor-steered decoder switch.

by the width of leg 1 to precisely the amount which leg 3 can accept. Therefore, when the setting pulse on leg 1 becomes larger than re-quired to transfer flux to leg 3, no further flux is available for transfer to leg 4 as is the case with a two-hole transfluxor. In this four-legged transfluxor, leg 2 is a dummy which always remains saturated in the same direction to provide the necessary return path to satisfy con-tinuity of flux flow.

[40] H. W. Abbott and J. J. Suran, "Multihole Ferrite Core Configurations and Applications," *Proc. I.R.E.*, Vol. 45, p. 1081, August, 1957.

[41] N. F. Lockhart, "Logic by Ordered Flux Changes in Multipath Cores," *I.R.E. National Convention Record*, March, 1958, Part 4, Vol. 6, p. 268.

[42] U. F. Gianola and T. H. Crowley, "Magnetic Networks for Performing Logic," *Proc. Special Technical Conference on Nonlinear Magnetics and Magnetic Amplifiers*, Los Angeles, Calif., p. 425, August 6-8, 1958. August 6-8, 1958.

[43] H. D. Crane, "A High-Speed Logic System Using Magnetic Elements and Connecting Wire Only," *Proc. Special Technical Conference on Non-linear Magnetics and Magnetic Amplifiers*, Los Ageles, Calif., p. 465, August 6-8, 1958.

The third aperture can also be used to reduce the time required to set the transfluxor. In general, with cores or transfluxors, it is found that the setting characteristic is the same no matter what the duration of the setting pulse, as long as it is longer than some minimum time, τ. This time τ might be defined as the nominal setting time of the transfluxor and is approximately equal to the switching time defined for optimum operation in a 2 to 1 regime for a memory toroid made of the same material. This switching time depends on the nature of the material and is typically between $1/2$ and 5 microseconds for good square-loop ferrites.

Fig. 23—Elimination of oversetting with an additional aperture.

It is found that the characteristic is quite different when the setting pulses are shorter than τ. For a given amplitude of setting pulse, much less flux is set than for the nominal or longer duration pulses. A typical curve for short setting pulses as well as the asymptotic characteristics for longer than τ times are shown in Figure 24. If many short setting pulses of a given amplitude are used, it is found that the setting will reach the value obtained with a single longer pulse of the same amplitude. In many applications where it is desired to sample the instantaneous value of an analog quantity and store it for further manipulation,

or simply store digital information, these effects are undesirable as is the case, for example, when it is desired to set a transfluxor by the coincidence of several short pulses applied to leg 1.

Rapid coincidence setting of a transfluxor is possible in spite of its inherent relatively slow setting characteristic.[38] The method is as follows: A d-c current is applied to leg 1 of the transfluxor in addition to the setting current coincident pulses. Consider the case, for example, of two such pulses, one of the pulses being purely for selection and the other carrying the analog or digital information to be stored.

Fig. 24—Setting of transfluxor by short pulses.

Let the amplitude of the first pulse and the maximum amplitude of the second be equal to the d-c bias. It is evident that the application of both pulses will set the transfluxor to the desired level in a desired time (if the amplitude is chosen to be large enough), whereas the application of either one separately will have no effect on the setting. It is interesting to note that after the termination of the setting pulses the d-c bias will restore leg 1 to full saturation, but this will not erase the setting, as the flux flow will be to the closer leg 2. The magnetic circuit through legs 1 and 2 "protects" the setting of the flux in the output legs 3 and 4.

It is also possible[44] to use three-apertured cores for high-speed current-coincident memories using a d-c bias winding in an arrangement somewhat similar to the one described above, but resulting in destructive rather than non-destructive read-out.

Transfluxors with three or more apertures can be made to perform various logic functions. Two simple principles are sufficient for the understanding of most effects in these more complicated transfluxors. These are:

(1) The total amount of flux through any plane section of a core will be zero to satisfy the condition of continuity of flux flow. Consequently, the transfer of flux from any branch of the core to other branches can be considered as lossless. The situation is analogous to the hydrostatics of an incompressible fluid, as the total amount of fluid remains constant no matter how it is distributed.

(2) The path followed in any flux change for a given magnetomotive excitation will tend to be as short as possible. If several alternate paths are offered to the flux flow, the flux will take the shortest path unless it includes a saturated leg in the direction of the prospective flow. In that case it will take the next shortest path. Following are a few illustrative examples:

A transfluxor with three apertures in a row, as shown in Figure 25, can be operated as a two-input sequential gate.[37] An output is produced if the two inputs A and B are applied to it in the order AB and no output is produced if either input is missing or if the two inputs are applied in the order BA. The operation is illustrated by the symbolic diagrams, Figures 25b, c and d.

The transfluxor with five apertures arranged in "flower" configuration (Figure 26) can be operated as a four-input "AND" gate.[37] The occurrence, in any order, of all four input signals A, B, C, and D is required to open the gate. The operation depends upon the fact that the output flux via legs 1, 2, 3, and 4 around the central aperture can be blocked by any one of the four legs, and is unblocked only when the senses of flux saturation around central hole in all legs are the same.

A transfluxor which can be set by either polarity of setting pulse, can be obtained by using four apertures.[37]

Another example of the multiple-hole transfluxor, is a six-hole "logicore" which can be used for a four-terminal odd-parity checker.[40]

The "Laddic"[42] is a multiaperture structure resembling a ladder,

[44] W. W. Lawrence, Jr., "Recent Developments in Very-High Speed Magnetic Storage Techniques," *Proc. Eastern Joint Computer Conference*, p. 101, December, 1956.

Fig. 25—Three-apertured transfluxor sequential gate.

Fig. 26—Five-apertured "flower" transfluxor.

with rungs and side rails as shown in Figure 27, with which it is possible to perform logic functions. It operates on the second principle mentioned above, namely that the flux change occurs on the shortest path available to it. A resetting pulse creates an initial flux configuration with alternate rungs oppositely magnetized (Figure 27). A subsequent driving current pulse on the end rung 1 tending to magnetize it downwards tends to magnetize upwards the flux in the nearest rung initially reset downwards. If large holding currents on input windings 2, 4 and 6 prevent any possible upward changes of flux in the corresponding rungs, the nearest rung capable of switching will be the end output rung 8. Therefore this rung will switch if, and only if, all three inhibiting inputs are present, while another rung will switch if any or all inputs are missing. The laddic is thus an "AND" gate. It is also possible to obtain "OR" and "NOT" functions. The holding currents must be larger than a certain minimum which is directly related to the amplitude of the clock pulse, but otherwise are not critical.

Fig. 27—Eight-rung laddic.

Multiple-hole cores or multiple-aperture devices, abbreviated MAD's[43] can be used for logic systems in which the elements are interconnected by electrical linkages made of wire only, and which do not include any isolating elements such as diodes or resistors. This can be illustrated by the example of a shift register made from MAD's, as shown in Figure 28. There are four "clock" windings: The clear E winding linking the center hole of all even-numbered cores, the clear O winding linking similarly the odd-numbered cores, the advance winding OE feeding the linkages between the transmitting leg T of an O core and the receiving hole R of an E core, and the advance winding EO feeding the similar linkages between even and odd cores. The clear E pulse leaves the E cores saturated clockwise. Let us assume that leg 3 of the O core to the left of a given E core is initially set, as shown in Figure 28. The effect of the next advance OE pulse will be to reverse the flux around the transmitting hole T of the O core (legs 3 and 4) and to reverse the flux along path 1-3 on the E core, leg 3 being the

closest leg magnetized in the proper direction. The effect of the subse-
quent clear O pulse will be to change the flux around path 2-4 of the O
core and the path 1-2 of the E core, leaving the state of legs 3 and 4
around the T hole of the E core unaltered. The subsequent EO and E
pulse have similar effects.

Two characteristics of the multi-aperture cores make the operation
possible: (1) The transmission of the setting from core to core is not
impeded by linkages of these cores to preceding and following cores in
the register because on the transmitting core the change of flux is local
around the T hole and thus has no effect on the linkage of the R hole
and because on the receiving core the flux change is of leg 3 and not
leg 4 linking to the next core. Similarly, the clearing operations have
no effect on the established settings. Diodes of a conventional core

Fig. 28—Shift register using multiple hole cores (MAD).

register having an isolation function are therefore not necessary. (2)
The transfer of flux from a core to another through a one-turn electrical
linkage is necessarily accompanied by some loss due to unavoidable
ohmic losses in the linkage. Therefore the amount of flux set in leg 3
of a receiving E core is less than the amount of flux change in leg 4 of
the transmitting O core. This loss of flux would cause a gradual
diminution of signal along the register if it were not compensated for
by an effective gain of flux within the E core during the subsequent ad-
vance pulse EO resulting from the fact that there is more flux available
for transfer to leg 4 than was originally set from leg 1 to leg 3. This
gain results from minor hysterisis loop properties of the core which
permit some flux from leg 1 to be transferred to leg 4 in addition to

that of leg 3 when a certain relatively high magnetomotive force is applied to leg 4. The amplitude of that magnetomotive force is so chosen that the additional flux is essentially zero when leg 3 is not set and yet is substantial when leg 3 is set. The operation is dependent on particular hysteretic properties which turn out to be enhanced by proper geometry of the core (not shown in the figure) and a relatively critical regulation of the amplitude of the advanced current pulses.

A variety of logic circuits[43] can be made from MAD's based on the principle of control of one core by another illustrated by the shift register.

CONCLUSIONS

The examples of devices and circuits described in this paper illustrate the variety of storing and switching tasks which can be accomplished by magnetics. Many of these tasks can be performed by other devices such as transistors and diodes, vacuum tubes, cryotrons, and other cryogenic devices, and by devices based on such physical phenomena as ferroelectricity, photoconductivity, electroluminescence, magneto-resistance, etc.

It is perhaps most significant to compare the art of magnetics to the art of semiconductor devices, as both are relatively well developed and both utilize solid-state devices at non-controlled ambient temperatures. In general, storage is better accomplished with magnetics since it is an inherent property of the materials used and does not depend on a flip-flop or other artifice required with semiconductor devices. High-speed selective access memories are likely to be magnetic in the foreseeable future as no other technique has been as successful to date. Plates and twistors are likely to succeed cores. Thin films also offer promise.

On the other hand, switching is better accomplished in many cases by semiconductor devices which have inherent gain and greater non-linearity. Switching by magnetics requires, in general, more power than by semiconductors and the power must be in the form of a-c or pulses, i.e., generated by tubes or transistors. In spite of these disadvantages, magnetic switching is preferable (1) when the number of switching channels is very large, as for switches to drive magnetic memories; (2) when the switching function is intimately related to a storing function, as in the transfluxor; (3) when extreme reliability is desired; and (4) for economy operation in special applications, particularly at lower speeds.

Magnetics is and is likely to remain a dominant tool for the manipulation of digital information.

UNIFIED REPRESENTATION OF JUNCTION
TRANSISTOR TRANSIENT RESPONSE*

By

A. HAREL AND J. F. CASHEN

RCA

RCA Industrial Electronic Products,
Camden, N. J.

Summary—*A unified analytical representation of junction-transistor small-signal transient response is developed, based on the assumptions that the frequency dependence of α is the dominating factor and that the collector load is relatively low. The result, which is expressed in a single simple equation, applies to any circuit configuration of the transistor, including circuits where any or none of the transistor terminals are grounded, and circuits with arbitrary source resistance. Various specific results, usually stated in separate expressions, can be derived from this general equation, such as effective cutoff frequency, rise-time, effective current amplification factor, response to current drive, etc. The emphasis is placed upon the quantitative effect which the external emitter and base resistors have on the performance of hybrid circuits. This theory is utilized in optimizing transient response by means of proper choice of external emitter and base resistors. The theoretical conclusions are verified by experiments.*

INTRODUCTION

THE use of transistors as linear pulse amplifiers requires an understanding of their transient behavior. Several techniques are available[1-3] for the transient analysis of transistors operating in the linear active region. However, it seems that they all lack a simple unified approach which expresses the collector transient response in a single mathematical formula. It is desirable to have a formula covering not only the grounded-base, grounded-emitter, and grounded-collector configurations, but also configurations containing external base and emitter resistors of arbitrary magnitude, and a signal source having an arbitrary internal resistance. The purpose of this paper is to present such a solution.

---

* Manuscript received October 15, 1958.

[1] R. F. Shea et al., *Principles of Transistor Circuits*, John Wiley & Sons, New York, N.Y., 1953.

[2] J. L. Moll, "Large Signal Transient Response of Junction Transistors," *Proc. I.R.E.*, Vol. 42, p. 1773, December, 1954.

[3] L. P. Hunter et al., *Handbook of Semiconductor Electronics*, McGraw-Hill Book Co., Inc., New York, N.Y., 1956.

As an example of the limitations of existing transient analyses, consider the following approach. Assume that

$$I_c = \alpha I_e \qquad \text{for grounded base,} \tag{1}$$

$$I_c = \beta I_b \qquad \text{for grounded emitter,} \tag{2}$$

$$\beta = \frac{\alpha}{1 - \alpha}, \tag{3}$$

and

$$\alpha = \frac{\alpha_0}{1 + \dfrac{s}{\omega_a}}, \tag{4}$$

where α_0 is the zero-frequency value of the grounded-base short-circuit current-gain factor, ω_a is the cutoff frequency of alpha, and s is a complex frequency. For a step of input current,

$$I_c = [\alpha_0 \, (1 - e^{-\omega_a t})] \, I_e \text{ for grounded base} \tag{5}$$

and

$$I_c = [\beta_0 \, (1 - e^{-\omega_\beta t})] \, I_b \text{ for grounded emitter} \tag{6}$$

where $\omega_\beta = (1 - \alpha_0) \omega_a$ is the cutoff frequency of beta, the grounded-emitter current-gain factor. The two collector-current expressions readily yield the familiar relations for 10-90 per cent rise time, t_r, and initial rate of rise of collector current.

$$t_r \cong \frac{2.2}{\omega_a} \tag{7}$$

and

$$\left. \frac{dI_c}{dt} \right|_{t=0} = \alpha_0 \omega_a I_e \tag{8}$$

$\left. \right\}$ for grounded base

$$t_r \cong \frac{2.2}{\omega_\beta} \tag{9}$$

and

$$\left. \frac{dI_c}{dt} \right|_{t=0} = \beta_0 \omega_\beta I_b = \alpha_0 \omega_a I_b \tag{10}$$

$\left. \right\}$ for grounded emitter.

This technique is easy to use, but it has obvious drawbacks. In

addition to the fact that separate equations are needed for each con-
figuration, this method applies only to idealized current drive. It does
not apply where sources with finite internal resistances are involved,
or where degeneration is employed. As shown in the following discus-
sion, all these cases can be described in a single, simple expression.

DERIVATION OF UNIFIED COLLECTOR CURRENT EQUATION

The unified collector-current equation is derived by examining the
various circuit configurations separately and showing that they lead
to an identical result. For the sake of generality, a voltage source, V_s,
with an internal series resistance, R_G, is used in all cases. The terms
"emitter input" and "base input" are used to distinguish the cases
with external series resistance from those where one terminal of the
transistor is grounded.

Fig. 1—Emitter-input circuit.

The emitter-input configuration is shown in Figure 1, where R_e'
denotes external resistance in the emitter lead, R_b' denotes external
resistance in the base lead, and R_L denotes load resistance in the col-
lector lead. In order to realize a simple, but still useful result, the
assumptions and equivalent circuit of Figure 2 are used where
$R_e = R_e' + r_e + R_G$, $R_b = R_b' + r_b$, and r_e, r_b, r_c, C_c are the internal
transistor parameters at the emitter, base and collector, respectively.
It should also be noted that the expression for the frequency depend-
ence of alpha, $\alpha = \alpha_0/(1 + s/\omega_a)$, includes no factor to account for
the initial delay in the collector response. The basic equations for the
circuit described in Figure 2 are

$$V_b' = V_s - I_e R_e = I_b R_b, \tag{11}$$

and

$$I_b = I_e - I_c = I_e (1 - \alpha). \tag{12}$$

Solving for I_e and multiplying by α,

Fig. 2—Emitter-input equivalent circuit.

$$I_c = \frac{\alpha V_s}{R_e + (1-\dot\alpha)\ R_b}.\qquad(13)$$

Consider now the equivalent circuit of the base input, collector output configuration. This is shown in Figure 3, where

$$V_b' = -I_e R_e = I_b R_b - V_s,\qquad(14)$$

and

$$I_e = I_b + I_c = I_b\ (\beta+1).\qquad(15)$$

Solving for I_b and multiplying by β,

$$I_c = \frac{V_s \beta}{R_b + (\beta+1)\ R_e}.\qquad(16)$$

Combining with Equation (3) gives

Fig. 3—Base-input, collector-output, equivalent circuit.

$$I_c = \frac{\alpha V_s}{R_e + (1-\alpha) R_b}. \tag{17}$$

Since this equivalent circuit also holds for the base-input, emitter-output circuit, Equation (17) also describes that case. Comparison of Equation (17) with Equation (13) shows that they are identical. The ramifications of this statement are discussed below.

The Laplace transform of Equation (17) is

$$I_c(s) = \frac{\dfrac{\alpha_0 V_s(s)}{R_e + R_b\,(1-\alpha_0)}}{1 + \dfrac{R_e + R_b}{R_e + R_b\,(1-\alpha_0)}\dfrac{s}{\omega_a}}. \tag{18}$$

It has thus been shown that all circuit configurations, whether base-input or emitter-input, with or without external base and emitter resistors, result in a collector response which is described by a single expression.

The inverse transform of Equation (18) for the case of a step of source voltage is

$$I_c = \frac{\alpha_0 V_s}{R_e + R_b\,(1-\alpha_0)}\,[1 - e^{-\omega_r t}] \tag{19}$$

where

$$\omega_r = \frac{R_e + R_b\,(1-\alpha_0)}{R_e + R_b}\,\omega_a, \tag{20}$$

and I_c and V_s are considered positive quantities. Equation (18) may be used for various input functions. For example, with a ramp function for the source,

$$I_c = \frac{\alpha_0 \dot{v}_s}{R_e + (1-\alpha_0) R_b}\left[t - \frac{1}{\omega_r}\,(1 - e^{-\omega_r t})\right] \tag{21}$$

where $\dot{v}_s$ expresses the rate of increase of source voltage per unit time.

As mentioned above, the various configurations have the same collector-current equation. The difference in the various circuits is in their input resistances. From the basic Equations (11), (12), (14), and (15),

$$r_{\text{in}} = R_c + (1 - \alpha) \, R_b \quad \text{for emitter input} \tag{22}$$

and

$$\boldsymbol{r}_{\text{in}} = R_b + \frac{R_c}{1 - \alpha} = R_b + (\beta + 1) \, R_c \quad \text{for base input.} \tag{23}$$

In the s domain,

$$r(s)_{\text{in}} = (R_c + R_b) \, \frac{s + \omega_r}{s + \omega_a} \quad \text{for emitter input} \tag{24}$$

and

$$r(s)_{\text{in}} = (R_c + R_b) \, \frac{s + \omega_r}{s + \omega_\beta} \quad \text{for base input.} \tag{25}$$

Derivation of Unified Emitter- and Base-Current Equations

Emitter Current as a Function of Time

The emitter current is also independent of configuration. Using the same equations employed to find input resistance, one obtains for a step of source voltage,

$$I_c = \frac{V_s}{R_c + (1 - \alpha_0) \, R_b} \left[1 - \left(\frac{R_b C_0}{R_c + R_b} \right) e^{-\omega_r t} \right]. \tag{26}$$

This equation shows an initial step in emitter current of magnitude

$$I_c = \frac{V_s}{R_c + R_b}. \tag{27}$$

In practice this is usually not realized.

Base Current as Function of Time

The base current as well as the collector and emitter currents are independent of the configuration. For a step function of source voltage,

$$I_b = \frac{V_s}{R_c + R_b} \left[1 - \frac{\alpha_0 R_c}{R_c + (1 - \alpha_0) \, R_b} (1 - e^{-\omega_r t}) \right]. \tag{28}$$

Thus the base current starts with a step of magnitude

$$I_{b0} = \frac{V_s}{R_c + R_b} \tag{29}$$

and decays with a time constant of $1/\omega_r$ to a final d-c value of

$$I_{b\infty} = \frac{V_s\,(1-\alpha_0)}{R_e + (1-\alpha_0)\,R_b}.\tag{30}$$

DISCUSSION OF UNIFIED REPRESENTATION

It has been shown above that the collector-current transient response to a step function of voltage applied between emitter and base is given by

$$I_c = \frac{V_s\alpha_0}{R_e + (1-\alpha_0)\,R_b}\left[1 - \exp\left(-\frac{R_e + (1-\alpha_0)\,R_b}{R_e + R_b}\,\omega_a t\right)\right].\tag{31}$$

If
$$\frac{V_s\alpha_0}{R_e + (1-\alpha_0)\,R_b} = I_r\tag{32}$$

and
$$\frac{R_e + (1-\alpha_0)\,R_b}{R_e + R_b}\,\omega_a = \omega_r = \frac{1}{T_r}\tag{33}$$

then
$$I_c = I_r\left(1 - e^{-\frac{t}{T_r}}\right),\tag{34}$$

where I_r represents the final d-c value of the resulting collector current, and T_r represents its time constant, both as functions of circuit resistances. If I_c and V_s are considered as absolute magnitudes, then Equation (31) is valid for any circuit configuration provided the collector load resistance is sufficiently low. This unified expression can be reduced to yield specific results, as shown below.

Cutoff Frequency

Cutoff frequency, ω_r—the angular frequency at which collector current is 3 decibels below its low-frequency value—is expressed in the general case by the unified expression of Equation (33) for arbitrary values of R_e and R_b. This reduces to the following forms. For grounded base,

$$R_b' = 0;\qquad R_e \gg R_b = r_b.\tag{35}$$

When substituted in Equation (33), this yields the familiar

$$\omega_r \cong \omega_a = \frac{1}{T_a}.\tag{36}$$

For grounded emitter,

$$R_e' = 0; \qquad r_c = R_c \ll (1 - \alpha_0) \ R_b. \tag{37}$$

When substituted in Equation (33), this yields the familiar

$$\omega_r \cong (1 - \alpha_0) \ \omega_a = \omega_\beta = \frac{1}{T_\beta}. \tag{38}$$

For base input with small emitter resistor,

$R_e' \neq 0; \qquad R_b \gg R_e$, but R_e not necessarily much smaller than

$$(1 - \alpha_0) \ R_b. \tag{39}$$

This yields, when substituted in Equation (33)

$$\omega_r \cong \left(1 - \alpha_0 + \frac{R_e}{R_b} \right) \ \omega_a \tag{40}$$

which, however, is only an approximation of the general expression.[2]

Response to a Step Function of Current

Equation (31), which shows the response of collector current to a step function of source voltage can also yield the collector response to a step function of current applied to emitter or base, and thus reduces to the familiar forms shown below. It is assumed that a current step can be represented by a limit of the ratio of an infinite voltage step to an infinite series resistance.

Using the conditions of Equation (35) and substituting

$$\frac{V_s}{R_e} = I_{es} \ (\text{step}) \tag{41}$$

in Equation (31), gives the familiar

$$I_c = I_{es} \ [\alpha_0 \ (1 - e^{-\omega_a t})] \tag{5}$$

for the grounded-base configuration.

If the conditions of Equation (37) and the substitution

$$\frac{V_s}{R_b} = I_{bs} \ (\text{step}) \tag{42}$$

are utilized in Equation (31), the familiar

$$I_c = I_{bs} \left[\beta_0 (1 - e^{-\omega_\beta t}) \right] \tag{6}$$

results for grounded-emitter configuration.

It should be noted that the response to an ideal current step is independent of the circuit resistors. Note also that the expressions included in the square brackets of Equations (5) and (6) represent the effective current amplification factors as functions of time.

Response Independent of Input-Terminal

The fact that the unified expression of Equation (31) is valid for both emitter input and base input can be interpreted as stating that the response of collector current is independent of the configuration, i.e., the ground point of the source voltage, so long as the resistors and the magnitude of the source voltage remain unchanged. It should be emphasized, however, that the input resistances, and hence the current and power drawn from the source, are different in the two cases. The Laplace transform of the resistance for each case is given in Equations (24) and (25). The initial value of the input resistance is identical in both configurations, namely,

$$r_{in} \Big|_{t=0} = R_e + R_b, \tag{43}$$

but the final value is

$$r_{in} \Big|_{t\to\infty} = R_e + (1 - \alpha_0) R_b \tag{44}$$

for emitter-input, and

$$r_{in} \Big|_{t\to\infty} = \frac{R_e + (1 - \alpha_0) R_b}{1 - \alpha_0} \tag{45}$$

for base input.

These results lead to the important conclusion that, under the operating conditions assumed in this analysis, base input is superior to emitter input provided both configurations include identical resistors and source voltage. Both configurations yield identical transient response while the base-input configurations requires less over-all source current (power).

Bandwidth Considerations

The approach used in this analysis leads to consideration of the transconductance–bandwidth product rather than the more usual gain–

bandwidth product. The transconductance, $G_m(s)$, obtained from Equation (18), is

$$G_m(s) = \frac{\alpha_0}{R_e + (1-\alpha_0)\,R_b}\; \frac{1}{1 + \dfrac{s}{\omega_r}}.$$ (46)

The term "bandwidth" applies to the transconductance. Due to the dependence of α on frequency, the transconductance is highest at low (zero) frequency and drops at higher frequencies. Hence the cutoff frequency, the frequency at which the transconductance drops 3 decibels below its low-frequency value, determines the bandwidth. Thus the bandwidth of the transconductance, and hence the bandwidth of collector current (for constant source voltage and fixed circuit resistors), is given by

$$B_m = \omega_r = \frac{R_e + (1-\alpha_0)\,R_b}{R_e + R_b}\,\omega_a.$$ (47)

The transconductance–bandwidth product, the product of the bandwidth (Equation (47)) and the zero-frequency transconductance

$$G_m = \frac{\alpha_0}{R_c + (1-\alpha_0)\,R_b},$$ (48)

is thus given by

$$G_m B_m = \frac{1}{R_e + R_b}\,\alpha_0\omega_a.$$ (49)

These considerations lead to some interesting conclusions. As seen from Equation (49) the gain–bandwidth product is essentially given by the product $\alpha_0\omega_a$, but is inversely proportional to the sum of emitter and base resistors. Hence, the way to increase this product is not only to increase $\alpha_0\omega_a$, but also to decrease the sum of the resistors as much as permissible under other restricting conditions.

The collector current, as a function of frequency, stays constant up to ω_r, independent of the configuration, and therefore a bandwidth of ω_a can be realized when the input signal is applied to the base, provided $R_e \gg R_b$. This mode of operation combines the advantage of grounded base, i.e., maximum bandwidth, with the advantage of grounded emitter, namely, high input resistance (at low frequencies) and low source power.

Initial Rate of Rise of Collector Current

From Equation (31) it follows that the initial rate of rise of collector current in response to a step of source voltage (compare Equations (8) and (10) above) is given by

$$\frac{dI_c}{dt}\Bigg|_{t=0} = \frac{V_s}{R_e + R_b}\,\alpha_0\omega_a\,. \tag{50}$$

This again is independent of configuration and is inversely proportional to the sum of emitter and base resistors as in the transconductance–bandwidth expression of Equation (49).

APPLICATION TO RISE-TIME AND CURRENT GAIN

Ten—Ninety Per Cent Rise Time

In comparison with Equations (7) and (9), Equation (31) yields a unified expression for 10—90 per cent rise time, t_r, of collector current, in response to a step of source voltage;

$$t_r \cong \frac{2.2}{\omega_r}\,, \tag{51}$$

which is independent of configuration.

Rise Time to Specified Collector Current

In lieu of the consideration of 10—90 per cent rise time, one may examine the time required for the collector current to rise from zero to a specified magnitude related to the input current. This leads to the study of current gain as a function of time.

Current Gain as a Function of Time

Current gain as a function of time is denoted by n, and expresses the instantaneous ratio of collector current to base current. This n should not be confused with β. Whereas β_0 is the d-c ratio of collector current to base current, β is defined as a function of (complex) frequency and is expressed by

$$\beta(s) = \frac{\beta_0}{1 + \dfrac{s}{\omega_\beta}} = \frac{I_c(s)}{I_b(s)}\,. \tag{52}$$

Its inverse transform into the time domain

$$\beta(t) = \beta_0 \omega_\beta e^{-\omega_\beta t} \tag{53}$$

is inherently different from

$$n(t) = \frac{I_c(t)}{I_b(t)}. \tag{54}$$

$\beta(t)$, the collector current in response to a unit impulse of base current, is not applicable in this analysis, whereas $n(t)$ may be applied as an interesting figure of merit and is further studied below.

Rise Time as Function of Current Gain

The term "rise time" as used below is defined as the time required for the collector current to rise to a value which yields a specified instantaneous current gain of n in response to a step function of the voltage source. The following paragraphs discuss the proper choice of the circuit resistors in order to minimize and evaluate this rise time.

Instantaneous Current Gain

The instantaneous current gain is a function of time, and varies (Equations (31) and (28))

from $n = 0$ at $t = 0$

to $n = \beta_0$ at $t = \infty.$ \tag{55}

However, at any intermediate time this current gain is also a function of R_e and R_b, and may be improved by proper choice of these resistors. The optimization of the instantaneous current gain can be examined in two ways:

(1) Proper choice of R_e and R_b in order to minimize the time which is required to achieve a pre-specified instantaneous current gain.

(2) Proper choice of R_e and R_b in order to maximize the instantaneous current gain which can be achieved at a pre-specified time.

Minimization of Rise Time

From Equations (31), (28) and (54), it follows that the time, t_n, required to achieve an instantaneous current gain of n, is given by

$$t_n = \frac{1}{\omega_r} \ln \frac{1}{1 - \dfrac{n}{\alpha_0} \dfrac{R_e + (1 - \alpha_0) R_b}{(n + 1) R_e + R_b}}. \tag{56}$$

For a given transistor (ω_a, α_0) and a specified instantaneous gain, n, the rise time, t_n, depends on R_e and R_b. It approaches its smallest value as $R_b \to 0$, namely,

$$t_{n(\min)} \to \frac{1}{\omega_a} \ln \frac{1}{1 - \dfrac{n}{\alpha_0} \dfrac{1}{n+1}} ; \qquad (57)$$

as $R_e \to 0$, t_n approaches its maximum value:

$$t_{n(\max)} \to \frac{1}{\omega_a} \frac{1}{1 - \alpha_0} \ln \frac{1}{1 - \dfrac{n}{\alpha_0} (1 - \alpha_0)} . \qquad (58)$$

Maximization of Instantaneous Current Gain

From Equations (31), (28), and (54), it follows that

$$n = \beta_0 \frac{1 - e^{-\omega_r t_n}}{1 + \dfrac{\beta_0 R_e}{R_e + R_b} e^{-\omega_r t_n}} . \qquad (59)$$

The highest instantaneous current gain, at a specified time, t_n, is approached as $R_b \to 0$

$$n_{(\max)} \to \beta_0 \frac{1 - e^{-\omega_a t_n}}{1 + \beta_0 e^{-\omega_a t_n}} ; \qquad (60)$$

the lowest gain, when $R_e \to 0$, is

$$\cdot \qquad n_{(\min)} \to \beta_0 (1 - e^{-\omega_\beta t_n}). \qquad (61)$$

Rise-Time and Impedance Matching

In some cases the base is connected to the voltage source through a transmission line, where it is required that there exist a perfect impedance matching at the transistor input in order to prevent any reflections. This means that the input resistance to the base must be equal to the characteristic impedance of the transmission line, Z_0, which is assumed to be purely resistive.

However, the step-function input resistance to the base is an exponential function of time. A perfect match is thus attainable only at

one instant, whereas there is an inherent mismatch at any other time, except in the following case. As Equation (28) shows, the time dependance can be completely eliminated if, and only if, $R_e \to 0$. Therefore there is only one way to realize a perfect impedance matching, namely to eliminate any resistance, R_c', in the emitter lead, and connect the transmission line to the base through a resistor R_b'' such that

$$R_e' = 0, \qquad Z_0 = R_b'' + r_b. \tag{62}$$

Fig. 4—I_c versus time for 2N139 transistor.

If an external emitter resistor must be included in the circuit because of other considerations, perfect impedance match can be achieved by paralleling the appropriate base resistor with an inductor which provides the proper time constant to completely compensate the internal time constant.

Experimental and Numerical Results

Figure 4 shows the calculated and measured curves of collector-current response to a step of source voltage. The theoretical curve is

ma

10.0

I_{c1} , I_{b1} for $R'_e = 0$, $R'_b = 2R_o$, $(R_e \approx 0$, $R_b = 1050\,\Omega)$

I_{c2} , I_{b2} for $R'_e = R'_b = R_o$ $(R_e = 470\,\Omega, R_b = 580\,\Omega)$

9.0

8.0

I_{c1}

7.0

6.0

CURRENT

5.0

4.0

3.0

2.0

I_{c2}

1.0

I_{b1}

I_{b2}

0

0   .1   .2   .3   .4   .5   .6   .7   .8   .9   1.0 μs

t

Fig. 5—Measured curves of I_c and I_b versus time.

calculated from Equation (31) and the experimental curves are meas-
ured in two configurations: base input and emitter input. The two
experimental curves practically coincide, as stated by the theory, and
are rather close to the theoretical curve, the error being less than 5 per
cent. This proves the soundness of the assumptions underlying the
theoretical analysis as well as the usefulness of the theory when ap-
plied to a hybrid circuit.

Figure 5 demontrates the current waveforms associated with the

<div align="center">*Table I*—Instantaneous Current Gain</div>

t_n, μs	n (min)		n ($R_e' = R_b'$)		n (max)
$\omega_a = 16.3 \times 10^6$ $\alpha_0 = 0.965$ $R_0 = 460\ \Omega$	$R_e' = 0$; $R_b' = 2R_0$ $R_e \approx 0$, $R_b = 1050\ \Omega$		$R_e' = R_b' = R_0$ $R_e = 470\ \Omega$, $R_b = 580\ \Omega$		$R_e' = 2R_0$, $R_b' = 0$
	Calculated (Eq. 61)	Measured	Calculated (Eq. 59)	Measured	Calculated (Eq. 60)
0	0	0	0	0	0
0.2	3.02	3.37	5.93	5.35	13.0
0.4	5.69	6.03	16.7	14.5	26.8
0.6	8.00	8.49	24.8	21.6	27.6
0.8	10.1	10.7	26.9	23.7	27.6
∞	27.6	27.6	27.6	27.6	27.6

problem of instantaneous current gain. Two configurations are compared: in one case R_e is minimized ($R_e' = 0$) and in the other case R_b is minimized ($R_b = R_e$) where the sum $R_e + R_b$ is equal in both cases. The instantaneous current gain, n, is always reached earlier if R_b, rather than R_e is minimized, as shown in Table I. In addition it may be seen that the initial base currents (disregarding their finite rise time, unavoidable in practical experiments) and initial slope of collector

a. Absolute Values   b. Relative Values

Fig. 6—Instantaneous current gain (calculated curves for normalized time-frequency scale; $\alpha_0 = 0.965$).

currents are approximately equal in both cases, as stated in the theoretical analysis.

In order to clarify the concept of the instantaneous current gain and its dependence on the emitter and base resistors, Figure 6 is included as a supplement to Figure 5 and Table I. This figure shows calculated curves for a normalized time frequency scale and $\alpha_0 = 0.965$. Both the absolute-value and the relative-value curves clearly demonstrate the difference in the rate of rise of the instantaneous current gain as a function of the circuit resistors.

THE MEGACODER — A HIGH-SPEED, LARGE-CAPACITY MICROMINIATURE DECODER FOR SELECTIVE COMMUNICATION*

By

H. KIHN AND W. E. BARNETTE

RCA Laboratories,
Princeton, N. J.

Summary—This paper describes an experimental microminiature digital decoder which can be preset to respond to any one of one million possible code combinations provided by a bipolar twenty-pulse binary code. This capacity can be extended indefinitely by the simple addition of identical microminiature modules. This Megacoder, occupying less than one cubic inch, has been incorporated in a pocket size FM personal paging receiver to provide a personal paging system capable of selectively calling one million subscribers at rates exceeding 5,000 calls per minute. Incorporated in the Megacoder is a unique magnetic-core transistor shift register element, selective combinations of which perform the above coding function with less than 70 milliwatts of power from the miniature receiver battery. Measurements have indicated reasonably error-free decoding performance at r-m-s carrier to fluctuation noise ratios of 15 decibels when operating from the miniature 150-megacycle FM receiver.

INTRODUCTION

THE need to extend mobile communications to personal paging without restricting the movements of the subscriber has become increasingly important in recent years. Applications of selective calling in hospitals, factories, aircraft communications and remote and automatic control of machine functions has underlined the need of a high-speed, high-capacity device of small dimensions, low power consumption, and high reliability. The limited availability of radio channels has emphasized the requirement of high speed and large subscriber capacity. These criteria are absolutely essential if a system be devised which is capable of future rapid growth as evidenced by the history of the telephone and mobile communications in general.

A personal paging or selective control system must provide a means for alerting one and only one of many subscribers or controlling a particular machine function. The limited radio-frequency space available for mobile communications prohibits the use of frequency division for selective calling. Therefore, methods have been devised through

---

* Paper received January 7, 1959.

the use of decoders for selectively transmitting calls to many individual subscribers over a single radio channel. A personal paging system utilizing a single narrow-band radio-frequency channel incorporates a radio receiver and decoder coupled together in a pocket-size carrying case. The receiver provides the function of detecting signals on the single channel. The decoder examines these incoming signals and activates an alarm when a transmitted code matches the code preset in the decoder.

This paper describes the development of a digital-type decoder which has a subscriber capacity exceeding one million and which is operable at calling speeds in excess of 5,000 calls per minute. The complete decoder occupies less than one cubic inch of space. This de-

INTELLIGENCE	SELECTION MEANS	MECHANIZATION
FREQUENCY BURSTS F_1 F_2 F_3 --- F_n	MECHANICAL RESONANCE	RESONANT REEDS
	ELECTRICAL RESONANCE	LC TUNED CIRCUIT
	ELECTRICAL RESONANCE	RC BRIDGE OR FEEDBACK
VARIABLE PULSE DELAY T_1 T_2 T_3 --- T_N	DELAY BETWEEN REFERENCE PULSE AND CODE PULSE.	ELECTRICAL ARTIFICIAL DELAY LINE OR ACOUSTIC DELAY LINE.
PULSE CODE MODULATION	DISCRETE COMBINATION OF POSITIVE AND NEGATIVE PULSES	TRANSISTOR- MAGNETIC CORE CIRCUIT. TRANSFLUXOR COINCIDENCE CIRCUIT. TRANSISTOR FLIP-FLOP CKT. FERRO ELECTRIC STORAGE CIRCUIT.

Fig. 1—Decoder systems.

coder as well as being very small has a capacity and coding speed many times greater than paging decoders now in use. It will be realized that expansion of this system to higher capacity is no problem since by adding one small module and increasing the volume by less than $1/40$ cubic inch the subscriber capacity is increased to over two million. Ultimate signaling speeds are affected mainly by available bandwidth and noise considerations.

DECODING SYSTEMS

Before adopting the digital system described here, several decoding systems were studied. An analysis of three modulation systems which could perform the decoding functions is shown in Figure 1. The intelligence to be transmitted is in column 1, the selection means in

column 2, and the mechanization or the circuitry of the function of column 2 is listed in column 3.

The frequency-burst method is most commonly used primarily because resonant reed decoders are old in the art and have been available as components for many years. This system involves the combinatorial use of m bursts of discrete frequencies F_1, F_2, F_3 - - - F_N, generally in the audio range and of sufficient duration to excite and cause to oscillate some type of mechanical or electrical resonant circuit. This modulation lends itself to normal voice-channel transmission, but means must be provided to insure that sustained voice tones or harmonic distortion do not actuate the decoder. The coding capacity or the number of possible distinct codes, C_f, depends on whether the resonant elements are actuated simultaneously or sequentially. In the latter, a storage element is required, such as a relay or RC circuit to carry over to the end of the transmission the information sent at the start. If N = total number of resonators and available (electrical or mechanical) and m units are used in each decoder, the coding capacity $C_f = N^m$.

For $N = 10$ and $m = 4$, $C_f = 10,000$

$N = 22$ and $m = 3$, $C_f = 10,648$.

In the simultaneous system,

$$C_f = {}_N C_m = \frac{N!}{m! \, (N-m)!}.$$

For $N = 10$ and $m = 4$, $C_f = 210$,

$N = 41$ and $m = 3$, $C_f = 10,660$.

Since the physical size of reed decoders militates against the use of more than 3 units, it is evident that for 10,000 codes, 22 resonators are required in a sequential system, whereas for the simpler simultaneous system, 41 frequencies must be available. The number of elements required is identical for mechanical or electrical resonant circuits, or for circuits where frequency selectivity is attained by RC negative feedback such as in a twin-T filter circuit.

For a system having identical time of resonator response throughout the available signaling band, both the resonance frequency, F_0, and the damping, or resonator Q_0, must be accurately controlled. Since the bandwidth for each resonator should be the same at all frequencies, the resonator Q must increase with frequency. It is shown in Ap-

pendix I that frequency-burst systems can be designed to provide the proper coding capacity and coding speed if component size and power were no limitation and mechanical resonator coupling to the electrical driving circuit could be made more efficient. This system is very attractive for slow coding and a system has been reported which can decode up to 15 codes per minute. However, the contact problem with resonant reeds, their susceptibility to mechanical actuation, physical size, and the inherent slowness of available resonators militate against their use in the subject decoder.

Other approaches to the problem are the use of LC filters and RC bridge or feedback circuitry. Since the frequencies must be precisely controlled, and since these are in the audio range, the components including delay means are too large indeed for personal paging use.

The second system involves the use of the variable delay between a reference pulse and a coding pulse, with a pulse-coincidence circuit as the correct-code indicator. Here combinations of T_1, T_2, T_3 - - - T_N are used to produce the required code capacity. The requirements for sequential and simultaneous operations are similar to those of the frequency-burst system, and include the need of storage in the sequential delay system. Since the coding delays are of the order of milliseconds, both lumped circuit and piezo-electric precision delay lines occupy too large a volume. Delay in solid-state devices, including transistors, depends both on amplitude and temperature, hence these are not yet suitable for large-capacity decoding

MEGACODER SYSTEM

A digital system using a binary code with pulse return to zero was adopted. Digital systems have a basic advantage in accuracy over analogue systems such as the two described above. The price paid for accuracy is the increased complexity and number of elements required to attain the desired coding capacity. The modulation consists of a discrete combination of positive and negative pulses which produce a coding capacity, $C_d = 2^N$ where N is the number of pulses. The transistor–magnetic-core mechanization was used for the Megacoder. The Transfluxor coincidence circuit required excessively large currents. The all-transistor register had excessive standby current and twice the number of transistors of the adopted system, and the ferroelectric storage system was eliminated because it required high operating voltages.

Figure 2 shows the modulation waveforms of two subscribers' codes in time sequence. Here a 20-digit bipolar binary code is used, giving a subscriber capacity of 2^{20} or 1,048,576 codes. A table is shown relating

subscriber capacity with the number of coding pulses from $N = 4$, equivalent to 16 subscribers, to $N = 24$, which allows 16,777,216 subscribers to use a single radio channel. The complete call consists of a coding interval using bandwidth-limited bipolar coding pulses and a readout or recognition pulse whose duration is approximately $\frac{1}{2}$ the total coding interval, or sufficiently long to allow separation of this pulse from an all-positive or all-negative pulse sequence by simple circuitry. The polarity of the recognition pulse may be either positive or negative so that its leading edge may serve as one of the coding digits as shown. The subscriber's call number may be assigned on a binary numbering system consisting of 5 groups of 4 pulses representing 2^0, 2^1, 2^2, 2^3, a positive pulse or ONE indicating the presence of a digit in the corresponding position and a negative pulse or ZERO, an

NUMBERS OF CODING PULSES n	SUBSCRIBER CAPACITY = 2^n
4	16
8	256
14	16,384
16	65,536
20	1,048,576
24	16,777,216

Fig. 2—Basic digital coding waveform.

absence. The arithmetic summation of the four pulse equivalent represents the number corresponding to the letter in the alphabet. Subscriber A's call could therefore be Jbgin and B's call, Ddame.

The use of both polarity pulses has the advantage of providing N shifting or clock pulses regardless of code, thus eliminating the need of an internal synchronized oscillator in each receiver. The transmitter thus effectively supplies the "clock" or synchronizing signals for thousands of receivers, thereby simplifying their design and reducing their cost.

A modification of the readout signal so as to more nearly equalize the net charge is shown in Figure 3a. This is a waveform of two test codes incorporating a bipolar readout pulse; the reverse excursion serves to compensate for the d-c component of the true readout pulse. This waveform requires more time per code and added circuitry to

allow the use of reversed polarity readout pulses. Both the single and bipolar readout pulses were used in system evaluation.

Figure 3b shows one modification of the bipolar readout pulse which allows the use of reversed polarity readout pulses. Here the equalizing pulse is serrated to distinguish it from the true readout pulse, allowing the use of both positive and negative readout pulses. A further modification involves the use of a variable duration equalizing pulse, as shown in Figure 3c, such that the net charge after each code approaches zero. This could be accomplished by averaging the positive and negative code pulses including the readout pulse during transmission, and using the resulting voltage to adjust the duration of the serrated equalizing pulse to produce a zero net charge. Since this involves only a low-pass filter and a delay multivibrator at the transmitter end of

Fig. 3—(a) Readout with equalizing pulse; (b) readout with serrated equalizing pulse; (c) readout with variable serrated equalizing pulse (zero net charge).

the system and has the added advantage of allowing complete freedom in code selection, it may advantageously be incorporated in any future selective call system. It should be observed that during system tests with repeated unidirectional codes at coding speeds of 2,000 or more codes per minute, no coding error was encountered, although the triggering or slicing level did vary somewhat. Undoubtedly, the presence of noise will require optimum slicing levels; hence, the noise performance should be improved by the use of the more complex equalizing readout waveform.

The decoder will operate with any type of coding pulse shape such as rectangular, triangular, sinusoidal, gaussian, etc. To conserve bandwidth, however, pulses approximating half cosine waves may be advantageously used. For a 3-kilocycle channel bandwidth, 1-millisecond

pulses will not produce sufficient undershoot to introduce coding errors. For the 20-element megacoder, 19-millisecond pulses and an 18-millisecond bipolar readout pulse will allow a coding speed

$$S_{BP} = \frac{60 \times 10^3}{t_p + t_r + t_c} = \frac{60 \times 1000}{37} = 1621 \text{ codes/minute}$$

where t_p = coding time, t_r = readout time, and t_c = compensation time. If a unipolar readout and 15 per cent pulse spacing are used instead of the above,

$$S_{V_1} = \frac{60 \times 10^3}{t_p + t_r + t_c} = 983 \text{ codes/minute},$$

For a variable-length serrated compensating pulse, S_V will range from S_{V_1}, for same polarity coding and readout pulses $(t_c = t_p + [t_r/0.636])$,

$$S_{V_1} = \frac{60 \times 10^3}{t_p + t_r + t_c} = 983 \text{ codes/minute},$$

to S_{V_2}, where the coding pulses and readout have opposite polarity $(t_c = 0)$,

$$S_{V_2} = 2140 \text{ codes/minute}.$$

If the bandwidth is unlimited, the decoding speed depends on the transistor alpha cutoff and the core switching time. Under these circumstances the components of the Megacoder allow

$$S_{V_1} = 9{,}830 \text{ codes/minute},$$

$$S_{V_2} = 21{,}400 \text{ codes/minute}.$$

It is evident that the system is capable of enormous signaling speeds. The use of higher-frequency transistors and ferromagnetic cores will allow proportional increase in coding speed.

Megacoder System Operation

Figure 4 shows a block diagram of the complete system. A 150-megacycle FM transmitter is modulated ±15 kilocycles by the digital modulation waveforms previously described and the signal is radiated from a centrally located elevated antenna. The receiver incorporates

Fig. 4—Selective call communication system.

both the decoder and a call alarm to indicate that the subscriber is paged.

In the Megacoder block diagram (Figure 5) the FM detector output is fed to a preamplifier which provides the necessary voltage levels in the transistor–magnetic-core circuitry and prevents pulse waveforms generated in the decoder from feeding back into the receiver. The positive and negative pulses are separated, and each serves to generate an extremely short positive pulse (3 microseconds). These pulses, rather than the millisecond-wide signaling pulses, are used throughout the decoder because of reduced battery drain and simplicity of providing necessary signal delays in the operating cycle. Additionally, the short pulses provide greater output from magnetic-core circuitry,

Fig. 5—Megacoder block diagram.

insure nondependence of decoder performance on received signal amplitude, and allow high signaling speed. This technique has been made feasible by the short switching time of the magnetic cores. The positive pulses are used as ONE's (information) and the summation of the negative and reversed pulses serve as negative unipolar shifting (trigger) pulses.

Both the trigger and information pulses are fed to a transistor–magnetic-core shift register which changes the magnetic states of miniature ferrite torroidal cores corresponding to the combinatorial arrangement of positive and negative pulses of the received signal. When the coding period is ended, the magnetic state of the cores is identical to the received signal pattern, regardless of previous signals or state of the register. The subscriber internal code consists of variation of the polarity of interconnection of readout windings on each of the twenty shift-register cores. The code may easily be changed by use of plug boards or by soldered interconnections. The readout pulse following the coding interval activates an interrogation circuit which effectively compares the magnetic state pattern stored in the register with the winding polarity pattern. If there is absolute coincidence in every core, which is the case for the correct code, a large current pulse flows in the circuit. If one or more cores have a magnetic polarity different from the readout winding polarity, the current is reduced. The received signal readout pulse is separated from the coding pulses in an integrator which drives a readout pulse generator connected in series with the readout windings of all the shift register cores. The generator pulse output current, whose amplitude is a function of the degree of coincidence between the signal code and internal code, is impressed on the "threshold amplifier." The latter serves to accentuate the difference between complete identity and near identity (one or two pulses different from the correct pattern). The output of the threshold amplifier drives a monostable multivibrator in whose output circuit is a fast-acting magnetic solenoid alarm device which attracts the subscriber's attention.

CODE-STORAGE SHIFT REGISTER

The shift register is considered first because the basic microminiature shift and storage element developed for this use also serves in both the input and readout circuitry.

The use of a core–diode shift register for storage of information in computers is well known.[1] The shortcomings of this method is the loss of power across each diode, since the diodes have high forward impedance at the small voltages involved. This results in a decreasing reliability of core setting with increasing number of stages. Further-

more, the transfer of state must be done by the trigger pulse which must be sufficiently large to insure complete transfer from the ONE to the ZERO state.

Figure 6 shows one stage of the shift register used in the Mega-coder. This encapsulated element .017 cubic inch in volume consists of an 80-mil ferrite core including five windings, a 2N105 transistor,

Fig. 6—(a) Megacoder shift-register stage; (b) Hysteresis characteristics of Megacoder ferrite core.

a resistor, and a capacitor. It also shows a typical hysteresis curve B versus H illustrating high remanance magnetization, and a curve of incremental permeability $\mu_i (= K\, dB/dH)$ versus coercive force, H. This latter characteristic is an important consideration in the method of information transfer and readout used in the Megacoder.

Figure 7 shows two stages of a shift register and oscillograms of pertinent electrical parameters. To illustrate the principle of oper-

ation, assume that a ONE had been set into the left hand core by a current pulse in the information winding. At this point of the magnetization curve, B_{R1} of Figure 6, the core permeability is very low; hence, the coupling between the collector and base winding of the transistor is insufficient to cause oscillation. When a trigger pulse, I_t, appears in the trigger winding, the magnetic operating point moves along the hysteresis curve so as to increase the permeability to the point where the feedback is sufficient to start oscillation. The collector current flows in such a direction as to reinforce the change of state begun by the small trigger pulse, and oscillation amplitude increases as the maximum permeability is traversed and finally ceases as the large transistor current, I_c, drives the core to extremely low coupling in the ZERO state. It is evident that I_c is a very narrow pulse of

Fig. 7—Shift register current and voltage waveforms.

approximately 3 microseconds duration. The transistor current during conduction is supplied mainly by the capacitor, and only to a small extent by the battery, thus almost completely discharging the capacitor as shown by V_c. At the cessation of transistor conduction, the capacitor is recharged through the resistor, which charging current, shown negative in Figure 7, serves to set the following core to a ONE state. The latter state is maintained until the incidence of the next trigger pulse at which time the same cycle is repeated for the next core, thus transferring information from core to core at each occurrence of a trigger pulse. If a core is originally in the ZERO state, no oscillation results upon being triggered. The lower oscillogram shows the waveforms for a 4-stage shift register whose output is fed back into the input to

produce cyclic information transfer. It is evident that transistor conduction and change of magnetic state of the core occur for every fourth trigger pulse. This arrangement was used very effectively in the course of the development of the Megacoder because the repetitive waveforms allowed observation and optimization of the circuit parameters. An equivalent circuit of the charge and discharge cycle is shown in Figure 8a; enlarged drawings of the time sequence of the trigger pulses, the condenser current and voltage for the discharge and reset cycle, and the effective impedance of the transistor are shown in Figure

(a)

(b)

Fig. 8—(a) Equivalent circuit of shift register stage;
(b) discharge and reset cycle.

8b. The transistor output resistance markedly affects the reset current since it does not become large until some time after capacitor current reversal, thus reducing the peak reset current and limiting the ultimate speed of shift-register operation. The 2N139 transistor is much improved in this respect but is physically larger than the 2N105.

The peak discharge current, i_d, is much greater than the peak charge current, i_c, because with the large feedback ratio used the transistor resistance drops to a few ohms during oscillation whereas the charging current flows through R_1 which is 200 ohms. Lower resistance cannot be used because of its damping effect during the oscillatory

cycle. The *RC* time constant is therefore optimized for a given core material and transistor type. The time constant used in the Megacoder register element is approximately 6 microseconds.

Figure 9 shows oscillograms of expanded views of the capacitor and transistor current pulses illustrating the transistor conduction "hangover," which limits the peak setting current and the ultimate speed of decoding. It also shows the relative decay, Δt, between core change of state and time of reset.

This type of shift register has proved very successful because of the small trigger drive required, extremely low power consumption, reliable setting and virtually no limitation as to the number of stages.

Fig. 9—Shift register current and voltage waveforms.

The latter advantage results from the power regeneration in each stage by means of one transistor rather than the use of two equally expensive diodes.

DETAILED MEGACODER CIRCUIT OPERATION

Figure 10 shows a schematic circuit diagram of the complete Megacoder arranged to correspond to the block diagram of Figure 5. The code signal amplifier, T_{r0}, incorporates negative feedback for temperature stability and an output filter, C_2, for optimizing signal-to-noise performance. The signal at the output of the amplifier is coupled through C_1 and is applied to the inputs of diodes X_1 and X_2 which are polarized so that X_1 conducts only on the positive signals and X_2 on the negative signals. The positive diode output, which is fed to the information pulse generator is differentiated by C_3 and the emitter

input impedance of T_{x_1} such that narrow positive pulses are applied at the emitter. The negative part of the differentiated pulse has no effect since the transistor conducts only during the positive pulses. These pulses cause T_{x_1} to conduct and in so doing produce regenerative coupling in core I which results in the generation of a 3-microsecond pulse in the collector circuit of T_{x_1}. This process completely discharges C_5 through transistor T_{x_1} and charges C_4 negatively with respect to ground. When the coupling in core I becomes essentially zero, causing cessation of T_{x_1} collector current, the recharging of C_5 through R_1 resets core I so that it is ready to receive the next positive code pulse.

Fig. 10—Megacoder schematic diagram.

Diode X_3 serves to discharge C_4 before the arrival of this next pulse. For coding speeds lower than 400 codes per minute, X_3 can be replaced by a resistor. The 3-microsecond pulse output of the information pulse generator is used to trigger the shift pulse generator through a winding on core II and, after a short delay provided by $R_D C_D$, sets a ONE into the first register core by means of the information driver stage shown as T_{x_3} and core III. By this process a positive input pulse first causes the register to shift its stored information and then sets a ONE into the first register stage.

Negative signals passing through diode X_2 are differentiated by C_6 and the parallel impedance of R_2 and the transistor base input resistance. The differentiated negative pulses then cause conduction in T_{x_2} and 3-microsecond shift pulses are regeneratively produced in the same manner as the information pulses described above. The shift pulse generator thus is triggered by both positive and negative signal

Fig. 11—Oscillograms showing Megacoder operation.

pulses and produces in its output twenty 3-microsecond pulses. These are fed in series with isolation diode X_4 and the shift pulse windings of the twenty-stage shift register. It is not necessary to clear the register before storing a new code because all information left in the register will be shifted out during the coding process. Figure 11a shows an oscillogram of the amplifier output for two different codes and 11b and 11c show the corresponding very narrow information and shift pulses.

By the action of the information and shift pulse generators, the twenty stages of the shift register assume magnetic states of ONES and ZEROS corresponding to the combination of positive and negative pulses of the incoming code. As mentioned previously, twenty code pulses are used regardless of the code combinations to allow the proper number of shift or clock pulses to be automatically generated in the decoder for each code received.

If the readout pulse following the coding interval is negative, which is the case for Figure 10, it is passed through diode X_2 and is coupled from the output of voltage divider R_3R_4 to the integrator circuit of the readout pulse generator which consists of R_5, C_8 and C_9. If a subscriber's code involved the use of a positive readout pulse, the integrator circuit would be fed from the output of the positive diode X_1. The readout pulse generator consists of the integrator circuit and a transistor–core regenerative circuit (T_{x_4}, core IV) similar to those in the information and shift pulse generators described previously. The integrator output waveform is shown in Figure 11d. When the integrator circuit voltage reaches a preset threshold level, the regenerative circuit produces a pulse which drives a pulse amplifier T_{x_5} through core V. The narrow readout current pulse at the output of the pulse amplifier is passed through isolation diode X_5 and into the series connected readout windings of the shift register. The extremely narrow pulse shape and the point of occurrence of this readout pulse is shown in Figure 11e. The high rate of rise of readout current, di/dt, results in greater resolution in this system which is inductance sensitive. Each of the readout windings is brought out to miniature terminals, and the polarity interconnection of these windings constitutes the subscriber's code. The comparison of the coercive force polarity and the magnetic state of each core due to the received code constitutes the logic of the system as described previously. The primary, P, of a very-square-loop, high coercive force core VII is connected in series with the shift register readout windings. If coincidence occurs due to the reception of a correct code, the magnitude of the resulting readout pulse current is sufficiently large to exceed the threshold of core VII, producing a narrow voltage pulse at its secondary, S. The secondary of this threshold amplifier core is connected to the quiescent emitter of a transistor monostable multivibrator, T_{x_7} and T_{x_8}. The collector circuit of T_{x_8} contains a fast-acting relay which actuates a spring-wound alarm. Therefore, with the reception of a correct code the alarm circuit is set off. The initial setting of the threshold core amplifier (VII) is accomplished by the inhibit amplifier. This amplifier is driven by core VI, which has its primary winding connected to the common

ground return of all the capacitors of the shift register. When any of the shift register stages trigger, the current through the primary shifts and resets the state of the inhibit amplifier core (VI). The voltage developed at the secondary of core VI drives the base of transistor T_{x_6}. The collector current output of T_{x_6} thus sets threshold amplifier core VII. This process serves to prevent unwanted readout activation of the alarm during the received code read-in period. This action also aids the recognition of a wrong code during the readout period in the following manner. If the readout pulse were to cause triggering of any of the register stages during the process of interrogation of a wrong code, the capacitor currents resulting therefrom would cause operation of the inhibit amplifier. This would prevent the

Fig. 12—Read-out integrator output for input coder with (a) eight positive pulses followed by eight negative pulses; (b) sixteen negative pulses; (c) eight negative pulses followed by eight positive pulses; (d) sixteen positive pulses.

transfer of pulse energy from primary to secondary windings of the threshold core. As a further aid to correct performance, the shift pulse generator is prevented from operating when the readout circuit triggers. The voltage for both T_{x_5} and T_{x_2} is supplied by a common capacitor, C_{10}, which is discharged when the readout circuit triggers. The reduction to zero of its collector voltage completely inactivates T_{x_2}. For the same reason the readout generator is immobilized during shift generator operation.

Figure 12 shows oscillograms of the readout integrator behavior for several representative codes including the extremes of sixteen negatives and sixteen positive pulses. The individual coding pulses appear at the left of each oscillogram, and if negative pulses are present, they are seen to cause a small rise in the integrated voltage. The true readout causes a large change in the integrated output which is enhanced by the operation of the trigger generator. This produces a regenerative

impulse shown as a sharp break in the exponential rise of voltage across the base of T_{x_4}. Additionally, this pulse places a charge on the emitter which keeps the transistor cut off for a time dependent on the emitter-circuit time constant. This action prevents double triggering during readout. The base-circuit discharge curve shows a cusp which represents the end of the readout interval, followed by the normal exponential decay.

It is evident that the coding pulses add a small amount to the integrator circuit charge, approaching an asymptotic value after about 8 pulses. Positive coding pulses are blocked by diode X_2 and hence have no effect on readout performance. Therefore, the only effect of the coding pulses on the readout threshold is to vary slightly the time at which the interrogation pulse is triggered. For a readout pulse of adequate duration this is of no consequence.

Fig. 13—Alarm multivibrator input. (a) Correct code; (b) incorrect code (1-bit error).

As mentioned previously, the output of the threshold amplifier is a narrow pulse whose magnitude depends on the degree of departure of the received code from the correct subscriber's code as represented by the interconnected readout windings. Such a pulse is shown in Figure 13, which is an oscillogram of the alarm multivibrator input which is fed from the threshold amplifier secondary winding. The larger amplitude pulse represents the output for a correct code and the short pulse for an incorrect code of 1-bit error. The amplitude ratio or "code resolution" is approximately 12 decibels. For greater divergence between correct and incorrect codes, this ratio increases correspondingly. For codes which result in triggering of the shift register transistors, the resolution is enhanced by the inhibition of the threshold amplifier operation as described previously. The major limitation to improved resolution is the core "squareness" as measured by the ratio B_r/B_s, where B_r is the remanent magnetization and B_s the saturation magnetization. However, a more appropriate comparison between

core materials should be on the basis of the dB/dH ratios, or incremental permeability, for the flat region and transition region of the hysteresis loop. When a correct code has been received, the pulse of Figure 13 is applied to the quiescent emitter of the monostable multi-vibrator T_{x_7}, T_{x_8}. As shown in Figure 10, the alarm solenoid is in the collector circuit of the normally cutoff transistor, T_{x_7}. The transition threshold is set by a bias E in the emitter of T_{x_8}, although a $-E$ voltage at the emitter of T_{x_7} would serve a similar purpose. Figure 11 shows the output waveforms of the alarm multivibrator for two different received codes; Figure 11a shows the code waveform. In Figure 11f, for code A correct, the multivibrator output pulse has time coincidence with this code only, and is zero for the B code. The reverse is true, as shown in Figure 11g, when the decoder is adjusted to respond to code B. These output pulses have approximately 15 milliseconds duration, but can be adjusted in width to suit the specific alarm requirements. The alarm solenoid resistance largely determines the peak current amplitude. For a 500-ohm solenoid, the peak current is 20 milliamperes.

MEGACODER PERFORMANCE

The Megacoder, both in breadboard form and in the microminiature package, was tested for accuracy of coding by the use of both punched tape and electronic code signal generators. Many different codes from all positive to all negative were used with negligible error. These tests included both a direct feed of code signals from signal generator to decoder and an over-all systems test involving an FM signal generator modulated by either the tape coder or an electronic code generator. The latter test may be well represented by the systems diagram of Figure 4 except for the substitution of a calibrated step-attenuator in place of the radio link. This allowed precise control of carrier-to-noise ratio (C/N) which is an important parameter in system performance The electronic code generator had the great advantage of permitting instantaneous selection of a wide range of digital codes. It has, however, been designed for only a 16-element system, hence many of the oscillograms which are included in this report were taken with a 16-element decoder driven by the electronic signal generator.

While the decoder system produces error-free operation when driven by a strong signal, the performance is markedly influenced by the presence of excessive noise, both fluctuation (white) noise and impulse noise. Figure 14 shows three oscillograms of coded signals with varying degrees of r-m-s carrier-to-noise ratios measured at the input to the decoder, which was fed by the 150-megacycle FM receiver. The use

of the stepped attenuator provided an excellent means of varying the fluctuation noise content of the signal. The decoder includes 4 decibels of additional filtering which can be increased for lower coding speeds. Optimum limiting has not yet been realized in this first model of the miniature receiver, hence noise performance might be expected to improve as development proceeds. The oscillograms of Figure 14 were translated into the curves of Figure 15 showing the probability of correct code reception versus carrier-to-noise ratio for a 16-bit code, a 20-bit code (Megacoder) and for a 20-bit code sent twice. The data for the 16-element decoder was obtained by connecting a counter to the alarm multivibrator output and recording the counts of proper alarm activation per fixed time interval for various signal-to-random-noise

a   C/N = 11 db   100% CORRECT

b   C/N = 6 db   66% CORRECT

c   C/N = -1.5 db   0% CORRECT

ADD 4 db TO C/N FOR FILTERED OUTPUT

Fig. 14—Discriminator output.

inputs. The signal source was the 150-megacycle signal generator modulated ± 15 kilocycles by the electronic digital-code generator.

It is evident that a sharp threshold exists for the satisfactory operation of the Megacoder in noise. For carrier-to-noise ratios greater than 15 decibels almost completely error-free operation results, while for 12 decibels, 1.7 per cent error, and for 8 decibels, 48 per cent error are indicated. It should be emphasized that these results are for the reception of a correct code only, so that an error represents a missed call and not necessarily an incorrect one. Missed calls arise from two sources. One occurs when the peak value of a noise pulse is large enough to depress a positive or negative code signal below the slicing level so that less than the required number of pulses are fed to the

decoder. This type of missed call requires the least amount of noise in the signal, since a peak value of noise equal to the slicing level (about ½ the peak value of the positive or negative pulse amplitude) is required. At higher noise levels, a pulse may be completely reversed, hence effectively changing the received code. The second factor is the readout coincidence circuit behavior in the presence of noise. Since the alarm actuation requires an AND coincidence, i.e., the reception and storage of a proper code AND the generation of a proper readout

Fig. 15—Effect of noise (r-m-s carrier-to-noise ratio in decibels) on correct code reception.

interrogation pulse, the absence of either one, or both results in a *missed call*. Since readout operation involves an integrating circuit, the small bandwidth of the latter (50 cycles) makes this more immune to noise than the code-storage circuitry which has a bandwidth dependent on the coding speed (approximately 1,000 cycles for 800 codes per minute). As described previously the passage of current through the shift-register charging capacitors inhibits the operation of the input to the alarm multivibrator. For low carrier-to-noise ratios the noise peaks will tend to cause shift-register operation during the readout interval, thus inhibiting alarm multivibrator operation. Since the

radio link is a frequency modulation system, the discriminator output
noise spectrum is triangular, i.e., the *higher frequency noise* compo-
nents predominate, hence with *noise alone* (in the absence of signals)
or for very weak signals the readout circuit will not receive adequate
low-frequency signals to generate an interrogation. This is one of
the more important factors in "loss of calls" at low carrier-to-noise
ratios, but serves a useful purpose in preventing false calls in the
absence of signal or in weak signal areas.

Increasing the coding capacity while maintaining constant code
speed increases the probability of coding error according to the
relationship

$$P_{n_2} = 1 - (1 - P_{n_1})^{n_2/n_1}$$

$$C_{n_2} = 1 - P_{n_2} = (1 - P_{n_1})^{n_2/n_1}$$

where $n_1 =$ bits in system $1 = (16)$,

$n_2 =$ bits in system $2 = (20)$,

$P_{n_1} =$ probability of error in system 1,

$P_{n_2} =$ probability of error in system 2,

$C_{n_2} =$ fraction of time correct code is received in system 2.

This is derived in Appendix II. The dashed curve in Figure 15 desig-
nated as "20-bit code (1)" resulted from the above relationship after
P_1 had been measured as described previously. The (1) refers to a
single call (no repetition) whereas the curve for "20-bit code (2)"
shows the improvement in accuracy of coding by a single repetition
of the call with some form of parity check. This is given by the
relationship

$$C_{n_2} = 1 - (P_{n_2})^2.$$

If we had m code repetitions as above, $C_{n_{2m}} = 1 - (P_{n_2})^m$ and the cer-
tainty of coding would be improved, but at considerable increase in
circuit complexity.

The foregoing discussion includes only the effects of fluctuation
noise on decoder performance. In determining decoder performance
under environmental conditions that might be experienced in large
cities, it is necessary to include the effects of impulse noise interference
and signal distortion due to multipath propagation.

Figure 16 is a photograph of the modularized microminiature
Megacoder which is less than 1 cubic inch in volume. The modules are
separated into two groups: the input–output group and the shift-
register group. The picture shows the input–output group arranged

Fig. 16—Modularized microminiature megacoder.

along the upper left side of the Megacoder and the twenty-element shift register along the lower right side. There are seven input–output modules of two different sizes: five are $\frac{1}{4} \times \frac{1}{2} \times \frac{3}{8}$ inch and two are $\frac{1}{8} \times \frac{1}{2} \times \frac{3}{8}$ inch. The components of each module are first assembled and then encased in araldite resin to form a shockproof, moisture-proof structure. Both groups are mounted on the printed board shown in Figure 17 which also shows the elements of the shift-register modules, and a four-stage assemblage of the modules. Figure 18 illustrates the manner in which the Megacoder has been incorporated into the 150-megacycle FM receiver. The receiver audio amplifier has been retained in this unit, although it would normally be replaced by a spring-wound alarm.

PRINTED BOARD

SHIFT REGISTER
MODULE
$\frac{1}{4} \times \frac{3}{16} \times \frac{3}{8}$
.0175 CU. IN.

ASSEMBLY OF
4 - MODULES

TRANSISTOR   CAPACITOR   RESISTOR   FERRITE CORE

Fig. 17—Megacoder components.

SELECTION OF OPTIMUM CODES

Although the Megacoder will operate satisfactorily with all codes such as combinations from all positive to all negative pulses, it is advantageous to eliminate several extreme codes which constitute less than one per cent of the total coding capacity. These are codes which either require more core squareness, closer voltage tolerances, increased power consumption or complicate the readout equalization problem of the decoder. Table I shows the effect on coding capacity of eliminating some of the more troublesome codes such as those containing all positive or all negative pulses, or a great preponderance of pulses of one polarity. The number of codes eliminated is $N = 2n!/(n-r)!r!$, where $n = 20$ and $r =$ number of pulses eliminated which varies from 0 to 5.

Elimination of 1.2 per cent of the codes as shown in row E of Table I greatly simplifies the decoder design yet still leaves well over a million codes.

Fig. 18—One million subscriber personal paging receiver.

CONCLUSION

A brief comparison of selective calling methods has been presented. These include systems using resonant vibrating reeds, resonant electrical circuits, bridge feedback frequency-selective amplifiers, variable-delay coincidence circuits, and pulse code modulation systems. In order to satisfy the requirement of high signaling speed, microminiature size, low voltage, low power consumption and to provide for a system capable of future growth, a new system of selective communication was conceived using a binary pulse code signaling waveform. A decoder of modular construction, occupying less than 1 cubic inch in volume and having a capacity of one million subscribers at signaling speeds greater than 5,000 calls per minute, has been built and demonstrated. This Megacoder incorporates a unique magnetic-core–transistor shift and storage element, selective combinations of which perform

Table I—Reduction in Coding Capacity by Elimination of Undesirable Codes for a 20-Bit Code

Type of Code Eliminated		Number of Codes $N_0 = 1,048,576$		Reduction in Coding Capacity
		Eliminated	Retained	
A	20+ 20−	2	1,048,574	.0002%
B	A and 19+ 19−	42	1,048,534	.004%
C	B and 18+ 18−	422	1,048,154	.04%
D	C and 17+ 17−	2,702	1,045,874	.27%
E	D and 16+ 16−	12,402	1,036,174	1.2%
F	E and 15+ 15−	43,402	1,005,172	4.1%

the above coding function at less than 70 milliwatts power drain. The decoder provides adequate power to operate a miniature spring-wound buzzer alarm which is part of the FM personal paging receiver. The noise performance of the Megacoder is comparable to other pulse code communication systems and the decoder may be operated in high noise environment by reduction of signaling speed and replacement of two modules which are capable of operating over a three to one speed range. Measurements have indicated reasonably error free performance at r-m-s carrier-to-fluctuation-noise ratios of 15 decibels when operating from the miniature 150-megacycle FM receiver. The decoder has been operated in the laboratory as part of a complete 150-megacycle FM paging system. The systems and techniques resulting from the Megacoder development are expected to contribute to more reliable selective communication between airborne or spaceborne vehicles and their ground stations as well as providing high speed personal signalling for large population centers. The small volume, weight and power requirements of the megacoder can make it useful in many other applications.

APPENDIX I—EVALUATION OF PARAMETERS IN A FREQUENCY-BURST
DECODER SYSTEM

Since resonator component size increases rapidly with decreasing
frequency, a lower limit of 400 cycles per second is assumed. An
upper limit of 2,500 cycles per second for the signaling bandwidths
will tend to minimize burst sidebands appearing outside the 3-kilocycle
limit.

If a coding speed $S_c = 240$ codes per minute is chosen, then the
time of response per resonator for an m resonator decoder is

$$T_s = 60/240 = \tfrac{1}{4} \text{ second for a simultaneous system,}$$

and $\qquad T_s = 1/(4m)$ second for a sequential system.

Assuming equal rise and decay time so that errors are not introduced
by a resonator not having recovered from a previous wrong code,

$$T_{s_1} = \text{rise time} = \tfrac{1}{2} \times \tfrac{1}{4} = \tfrac{1}{8} \text{ second (simultaneous)},$$

$$T_{s_1} = 1/(8m) \text{ second (sequential)}.$$

Bandwidth $= \Delta F_{s_1} = .7/T_{s_1} = 5.6$ cycles per second.

If a single tuned resonator and a 12 decibel attenuation between desired
and undesired frequencies is required, a universal resonance curve
shows that the total channel width required is approximately $4\Delta F$.

$$\Delta F'_{s_1} = 22.4 \text{ cycles per second (simultaneous)},$$

and $\qquad \Delta F'_{s_1} = 22.4m$ cycles per second (sequential).

For a signaling bandwidth of 2,400 cycles, the number of resonators
which can be accommodated is

$$N_{s_1} = 2,400/22.4 = 107.$$

The resonator Q's will range from $400/5.6 = 71.5$ to $2,800/5.6 = 500$.

Such values of Q are difficult to attain by audio electrical circuits,
but can be attained with reed or piezoelectrical resonators. For sequen-
tial systems the range of Q will be $71.5/m$ to $500/m$.

This is an idealized case with no provisions made for delays in cir-
cuit or contact activation. It serves to indicate that if no size limitations

existed and if mechanical vibrating reeds could be designed with the proper Q, large-capacity, high-speed systems would be possible. However, due to the difficulty of coupling and the power needed to maintain mechanical oscillation, mechanical resonators are usually made extremely high Q which seriously limits the speed of response.

It is evident that for a given signaling bandwidth, a sequential system has larger capacity, but lower speed than a simultaneous system. If for the simultaneous system requiring 41 resonators the 2,400 cycles per second band is divided into equal increments, and a $4\Delta F$ guard band is assumed as before,

$$\Delta F_s = 2{,}400/4 \times 41 = 14.6 \text{ cycles per second,}$$

$$Q \text{ at } 400 = 400/14.6 = 27.4,$$

$$Q \text{ at } 2{,}800 = 2{,}800/14.6 = 191.5,$$

$$S_{c_1} = \text{coding speed} = 60/(2T_1) = 30\,\Delta F_s/.7 = 30 \times 14.6/.7$$
$$= 375 \text{ codes per minute.}$$

for zero code separation; for code separation equal to code length,

$$S_{c_2} = \tfrac{1}{2} \times 375 = 187 \text{ codes per minute.}$$

Appendix II—Coding Error Probability

$$P_{n_1} = 1 - (1 - P_0)^{n_1},$$

$$(1 - P_0) = (1 - P_{n_1})^{1/n_1},$$

$$P_{n_2} = 1 - (1 - P_0)^{n_2},$$

$$P_{n_2} = 1 - (1 - P_{n_1})^{n_2/n_1},$$

$$C_{n_2} = 1 - P_{n_2} = (1 - P_{n_1})^{n_2/n_1}.$$

For $n_1 = 16$ and $n_2 = 20$,

$$C_{n_2} = (1 - P_{n_1})^{1.25}.$$

where $n_1 =$ number of bits in system 1,

$n_2 =$ number of bits in system 2,

$P_0 =$ probability of error for 1-bit code,

$P_{n_1} =$ probability of error for n_1-bit code,

$P_{n_2} =$ probability of error for n_2-bit code,

$C_{n_2} =$ fraction of time correct code is received for n_2-bit code.

RCA TECHNICAL PAPERS†

Fourth Quarter, 1958

Any request for copies of papers listed herein should be addressed to the publication to which credited.

"Power Supply and Hi-Voltage Modifications Stressed in Latest Color TV Receivers," W. H. Fulroth, *Service* (October) 1958

"PPI Scan Generators for Use with Display Storage Tubes," *RCA Application Note AN-177*, Electron Tube Division, Radio Corporation of America, Harrison, N. J. (December) 1958

"The Prediction of Derivatives of Polynomial Signals in Additive Stationary Noise," I. Kanter, *I.R.E. Wescon Convention Record, Part 4, Automatic Control, Electronic Computers Human Factors in Electronics, Information Theory* 1958

"Publications and the Project Organization," M. H. Lowe, *I.R.E. Wescon Convention Record, Part 9, Engineering Writing and Speech, Engineering Management* 1958

"Pulse Amplifier with Submillimicrosecond Rise Time," F. Sterzer, *Rev. Sci. Instr.* (December) 1958

"On the Quality of Color-Television Images and the Perception of Color Detail," O. H. Schade, Sr., *Jour. S.M.P.T.E.* (December) 1958

"On the Quality of Color-Television Images and the Perception of Color Detail," O. H. Schade, Sr., *RCA Review* (December) .. 1958

"Quantum Limitations in Vision and Television," A. Rose, *Acta Electronica* 1957-1958

"Quantum Limitations in Vision and Television," A. Rose, *Optica Acta* 1958

"Radiative Surface Effect in Germanium," J. I. Pankove, *International Journal of the Physics and Chemistry of Solids*, Pergamon Press, Inc. 1958

"The RCA Flight Data System," C. N. Batsel, Jr., R. E. Montijo, Jr., and E. J. Smuckler, *I.R.E. Convention Record, Part 5, Aeronautical and Navigational Electronics, Instrumentation, Telemetry and Remote Control* 1958

"The RCA Test Facilities for TV Antennas," H. E. Gihring, *Broadcast News* (October) 1958

"Recent Work on Photoemission and Dark Emission Problems," R. W. Engstrom, R. G. Stoudenheimer, H. L. Palmer and D. A. Bly, *Trans. I.R.E. PGNS* (December) 1958

"Recommended Tube Types for Amateur Short-Wave Receivers," L. Aurick and P. Boivin, *QST* (November) 1958

"Reduction of Bandwidth Requirements for Radio Relay Systems," A. Mack, A. Meyerhoff, and Co-authors, *I.R.E. Convention Record, Part 8, Communications Systems, Military Electronics, Radio Frequency Interference, Vehicular Communications* ... 1958

"Remote Control of the 50-KW AM 'Ampliphase' Transmitter," C. J. Starner, *Broadcast News* (October) 1958

"Semiconductor Surface Potential and Surface States from Field-Induced Changes in Surface Recombination," G. C. Dousmanis, *Phys. Rev.* (October 15) 1958

"The Space Environment — A Preliminary Study," R. A. DiTaranto and J. J. Lamb, *Electrical Manufacturing* (October) 1958

"Status of Multiplier-Phototube Development for Scintillation Counters," W. Widmaier, *Trans. I.R.E. PGNS* (December) 1958

"Stereophonic Amplifier for Home Phonographs," R. S. Fine, *Electronics* (December) 1958

"Talos Land Based System Digital Checkout Equipment," F. X. Beck, *I.R.E. Wescon Convention Record, Part 8, Communication Systems, Military Electronics, Vehicular Communications* ... 1958

"Television Control Room Human Engineering Problems," E. B. Pores, *Jour. S.M.P.T.E.* (October) 1958

"Test Equipment for Color Servicing Varies with Needs of Service Man," J. R. Meagher, *Service* (October) 1958

AUTHORS

WILLIAM E. BARNETTE received the B.E.E. degree in 1953 from the University of Delaware. In June of 1953 he joined RCA Laboratories. In 1954, he was assigned to work in the Systems Research Laboratory, where he engaged in developing linear and logarithmic i-f amplifiers. Currently he is doing research in pulse code, digital communications and computer systems. Mr. Barnette is a Member of the Institute of Radio Engineers.

J. F. CASHEN received the B.S. degree in Electrical Engineering from Drexel Institute of Technology in 1957 and is currently working for the M.S. degree at the University of Pennsylvania. He joined RCA in 1957 as a member of Control Systems Development, Electronic Data Processing Division, Camden, N. J. He has been engaged in development of digital circuits for a pipeline control system, and circuit development for a high-speed data processing system. Mr. Cashen is a Member of the Institute of Radio Engineers.

HENRY B. DeVORE received the B.S. degree in physics in 1926 and the M.S. degree in 1927 from Pennsylvania State College and the Ph.D. degree in 1934 from the California Institute of Technology. From 1927 to 1931, he was employed at the DuPont Experimental Station. He joined the research division of RCA Manufacturing Co. at Harrison, N. J. in 1934, transferring to RCA Laboratories Division at Princeton, N. J., in 1942, where he worked on the development of television pick-up tubes and on microwave antenna problems. From 1945 to 1947, he was associated with Remington Rand, Inc. Since 1947, he has been at RCA Laboratories, Princeton, N. J., engaged in fundamental research on solids. Dr. DeVore is a Member of Sigma Xi and the American Physical Society and a Senior Member of the Institute of Radio Engineers.

ARIE L. EICHENBAUM received the B.Sc. and M.Sc. degrees in Electrical Engineering from Polytechnic Institute of Brooklyn in 1953 and 1954, respectively. From 1953-1954 he held a research fellowship at the Microwave Research Institute of Polytechnic Institute of Brooklyn. He joined RCA Laboratories as a member of the Technical Staff in June, 1954, where he has been engaged in research in the field of traveling-wave tubes and electron guns. Mr. Eichenbaum is a Member of the Institute of Radio Engineers.

185

ABRAHAM HAREL received the B.S. degree in Electrical Engineering from the Israel Institute of Technology in 1952, and the Dipl. Ing. degree from the same institute in 1953. In 1956 he received the M.S. degree from the University of Pennsylvania, where he was a research assistant for two years. He received the Dr. techn. sc. degree from the Swiss Federal Institute of Technology in 1958, while working with Zellweger Ltd., Switzerland. He joined RCA in 1953 and spent a year in the Camden, N. J. area as a trainee under the Foreign Exchange Visitor Program. He returned to RCA in 1958 and is currently working at the Control Systems Development Section, Electronic Data Processing Division. His work has mainly been in the fields of transistor switching circuits and digital systems. Dr. Harel is a Member of Sigma Xi and the Institute of Radio Engineers.

HARRY KIHN received the B.S. degree in Electrical Engineering from the Cooper Union Institute of Technology in 1934 and the M.S. degree from the University of Pennsylvania in 1952. Prior to joining RCA he was employed by the Hygrade-Sylvania Co., Polytherm (his own company), and the Ferris Instrument Company. He joined the RCA Manufacturing Company in Camden in 1939 and engaged in television field testing, receiver circuit development, and research in frequency modulation, fluctuating noise, ferromagnetic materials, and the APN-1 radio altimeter. In 1942, he was transferred to RCA Laboratories in Princeton, where he engaged in research on X-band receivers, silicon crystals, frequency modulation radar, automatic bombing devices, and the APN-22 altimeter. With the advent of color television, he participated in the development and design of the field test receivers. In 1953, he transferred to the System Research Laboratory, where he is doing research on millimeter wave high resolution radar. He is now directing research in pulse code and digital communication and space systems. Mr. Kihn is a Senior Member of the Institute of Radio Engineers, Sigma Xi, and the Operations Research Society of America.

MURRAY LAMPERT received the B.S. degree in 1942 and the M.S. degree in 1945 from Harvard University. During World War II he spent two years as an instructor at the Harvard Electronics Training School for the Armed Services, and one year in charge of the optical design section of the Harvard Observatory Optical Research Project. From 1945-1948 he was in the theoretical section of the Radiation Laboratory of the University of California engaged in study of high-energy physics. From 1949-1952 he was in the vacuum-tube section of the Federal Telecommunications Laboratories working on control of microwave energy by gaseous discharges. He was a co-inventor of the first gyrator using a gas tube. Since 1952 he has been with RCA Laboratories, Princeton, N. J.; he is presently a member of a group conducting theoretical studies of solid-state problems.

LEON S. NERGAARD received the B.S. degree in Electrical Engineering from the University of Minnesota in 1927, the M.S. degree from Union College in 1930, and the Ph.D. degree from the University of Minnesota in 1935. From 1927 to 1930, he was in the research laboratory and vacuum-tube engineering department of the General Electric Company; from 1930 to 1933 a teaching assistant in the department of physics at the University of Minnesota; from 1933 to 1942 in the research and development laboratory of the RCA Manufacturing Company; and since 1942 at RCA Laboratories in Princetno, N. J. Dr. Nergaard is a Member of Sigma Xi and the American Association for the Advancement of Science, and a Fellow of the Institute of Radio Engineers and the American Physical Society.

ROLF W. PETER received the M.S. degree in Electrical Engineering in 1944, and the Ph.D. degree in Radio Engineering in 1948 from the Swiss Federal Institute of Technology in Zurich, Switzerland. From 1946 to 1948 he was Assistant Professor of Radio Engineering at the Swiss Federal Institute of Technology. In 1948 he joined RCA Laboratories in Princeton, N. J. He was first engaged in traveling-wave amplifier research. From 1953 until 1957, he supervised the microwave and gaseous electronics research group. Since 1957 he has been Director of the Physical and Chemical Research Laboratory. Dr. Peter is a Member of Sigma Xi and of the American Physical Society, and a Fellow of the Institute of Radio Engineers.

JAN A. RAJCHMAN received the diploma of Electrical Engineering in 1934 and the degree of Doctor in Technical Sciences in 1938 from the Swiss Federal Institute of Technology. He joined RCA in the summer of 1935, and has been engaged in research since 1936, first in Camden and since 1942 in Princeton at RCA Laboratories. He is now Associate Director, Systems Research Laboratory and is directing research in electronic digital computers and data processing, a field in which he has been active since 1939. Dr. Rajchman is a Fellow of the Institute of Radio Engineers, a member of the American Physical Society, Sigma Xi, and the Association for Computing Machinery, and a co-recipient of the Levy Medal of the Franklin Institute—1946.

ALBERT ROSE received the A.B. degree from Cornell University in 1931 and the Ph.D. degree in Physics in 1935. From 1931 to 1934 he was a teaching assistant at Cornell University. From 1935-1942 he was with the RCA Tube Division in Harrison, N. J., and since 1942 he has been associated with RCA Laboratories at Princeton, N. J. From 1955 to 1957 he directed the research at Laboratories RCA, Ltd., in Zurich, Switzerland. Dr. Rose is a Fellow of the Institute of Radio Engineers and of the American Physical Society.

ROLAND W. SMITH received the B.S. degree from Western
Kentucky State College in 1939, and the M.S. degree in
Physics from Northwestern University in 1942. He was
with the Central Scientific Company in Chicago during
the period 1939-1940. From 1942 to 1947 he did research
in photoconductors at Northwestern for the Office of
Scientific Research and Development. In 1947 he joined
the Research Department of RCA Laboratories Division
at Princeton, N. J. Mr. Smith is a Member of the Amer-
ican Physical Society and Sigma Xi.